ESC-TR-97-065

Architecture and Concept of Operations for a Warfighter's Internet

Volume 2: Appendices

Prepared for:
Defense Advanced Research Projects Agency
Information Systems Office

Edited by:
Massachusetts Institute of Technology
Lincoln Laboratory

28 January 1998

Prepared for the Defense Advanced Research Projects Agency under
Air Force Contract F19628-95-C-0002

This report is based on studies performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology. The work was sponsored by the Defense Advanced Research Projects Agency, ISO, under Air Force contract F19628-95-C-0002.

The ESC Public Affairs Office has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

Gary Tutungian

Gary Tutungian
Administrative Contracting Officer
Contracted Support Management

Architecture and Concept of Operations for a Warfighter's Internet

Volume 2: Appendices

Prepared with Contributions from:

Air Force Rome Laboratory
Army CECOM
Jet Propulsion Laboratory
MIT Lincoln Laboratory
MITRE Corporation
Navy Command Control Ocean Surveillance Center
Naval Research Laboratory

Prepared for:

Defense Advanced Research Projects Agency,
Information Systems Office

28 JANUARY 1998

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

LIST OF ILLUSTRATIONS

Figure No. **Page**

Appendix A

Appendix B

Appendix C

LIST OF ILLUSTRATIONS (Continued)

Figure No. Page

Appendix D

Appendix E

Appendix F

LIST OF ILLUSTRATIONS (Continued)

Figure No. Page

Appendix H

Appendix I

Appendix J

Appendix K

LIST OF ILLUSTRATIONS (Continued)

Figure No. **Page**

Appendix L

LIST OF TABLES

Table No. **Page**

Appendix K

APPENDIX A

CROSS-LINKS SEGMENT DESCRIPTION (ARCHITECTURE OPTION 2)

The Warfighter's Internet (WI) is envisioned to bring multimedia, multicast, and Quality of Service (QoS) networking capabilities to the warfighter. These include voice, video, web browsers, white boards, global data base access, fire control, etc. The proposed architecture of WI resembles a cellular/PCS (Personal Communication Service) system with base stations that reside on airborne platforms, such as Unmanned Airborne Vehicles (UAVs). Access to the WI base stations is through small, handheld transceivers or via traffic concentrators.

Unlike commercial, cellular/PCS systems, the WI base stations are mobile. Consequently, one of the key technical issues to be solved is how to maintain communication among all the base stations of the WI. A multi-channel wide area subnet is proposed as the method for doing this. This section and the following one provide an introduction to a preliminary architecture and operational concept for the proposed subnet, which we shall call the Cross-Links Subnet. Many details remain to be worked out. Some of these details depend greatly on the capabilities of the RF hardware, modems, and encryption devices, which are not fully known at this time.

1. Cross-Links Subnet Architecture

This section describes the architecture of the Cross-Links Subnet in terms of the nodes that make up the subnet, the kinds of links employed, and the configuration of the Cross-Links Subnet nodes.

Figure 1 shows a possible scenario for the WI. In this figure, the nodes and links that make up the Cross-Links Subnet are shown in green. Note that, in addition to the airborne nodes, which contain the cellular/PCS base stations, the Cross-Links Subnet may contain other nodes. The latter are generally nodes that require high-data-rate links to the airborne nodes. One such node is the platform on which the command center is located. In any event, it is expected that the Cross-Links Subnet will contain a modest number of nodes (probably in the range 2 to 30). The links of the Cross-Links Subnet are of two types: high-data-rate links that are supported by directional antennas and lower data rate links that are supported by omnidirectional antennas.

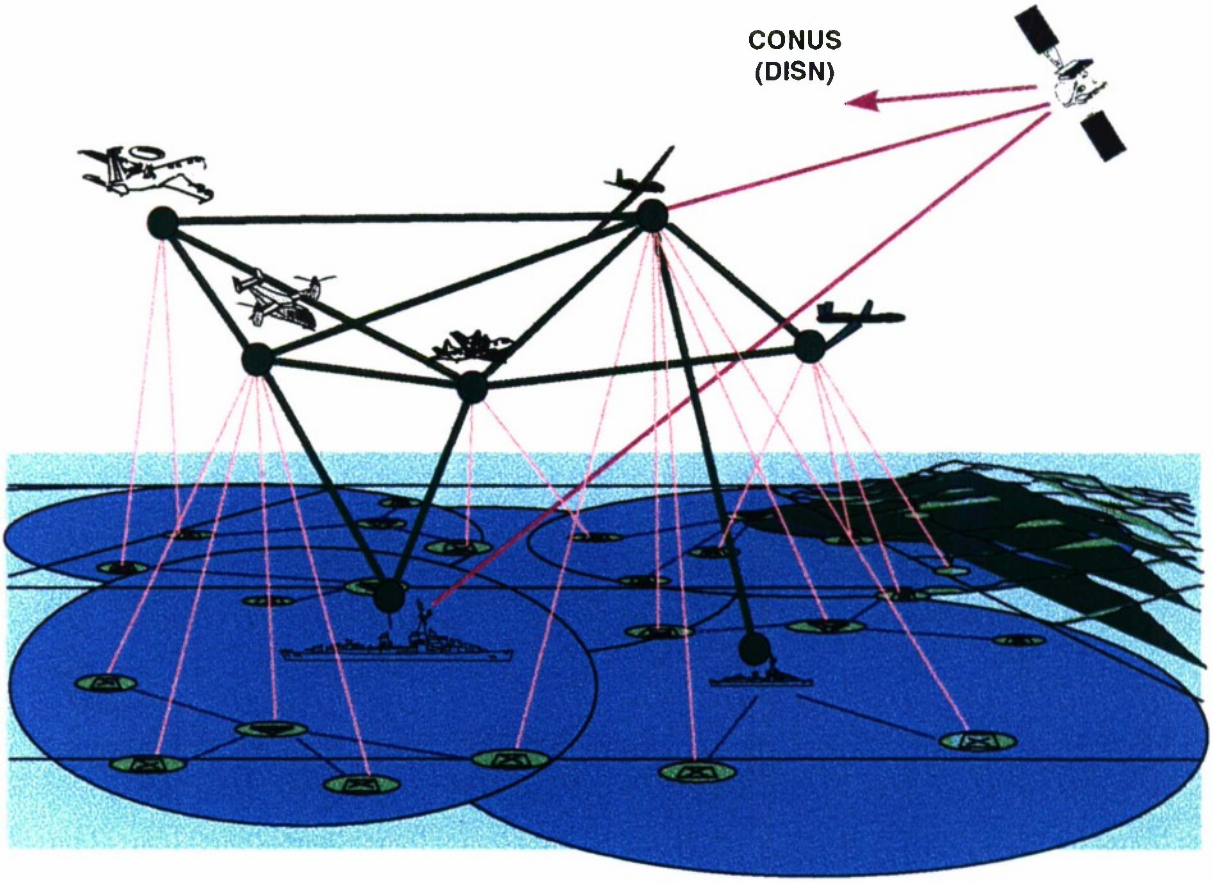


Figure 1. Possible scenario showing the Cross-Links Subnet (green) portion of the WI.

Figure 2 shows several possible configurations for platforms that house Cross-Links Subnet nodes. In each configuration, the Cross-Links Subnet node has at least one transmitter that uses an omnidirectional antenna. Each of these nodes is preassigned a unique frequency-hopping (FH) code for this transmitter. In order to receive the omnidirectional transmissions from several other Cross-Links Subnet nodes simultaneously, each node has a bank of receivers. The receivers may share a common, omnidirectional antenna. The transmit and receive antennas are positioned to minimize self-interference, and the FH patterns are designed to reduce self-interference, so as to permit simultaneous transmit and receive operation. In the ideal case, if there are N nodes in the Cross-Links Subnet, each node would have N-1 receivers available for listening to the omnidirectional transmissions. If there are less than N-1 receivers at a node, one of these receivers should be used for scanning in order to ascertain the best set of nodes to listen to with the remaining receivers. Some of the Cross-Links Subnet nodes, such as the one on the command center's platform and those on UAVs, also contain several directional antennas and associated transmit and receive strings. These are used to set up high-data-rate links between select Cross-Links Subnet nodes. Selecting and setting up the high-data-rate links is normally done automatically by the Cross-Links Subnet; however, it may also be done manually if desired.

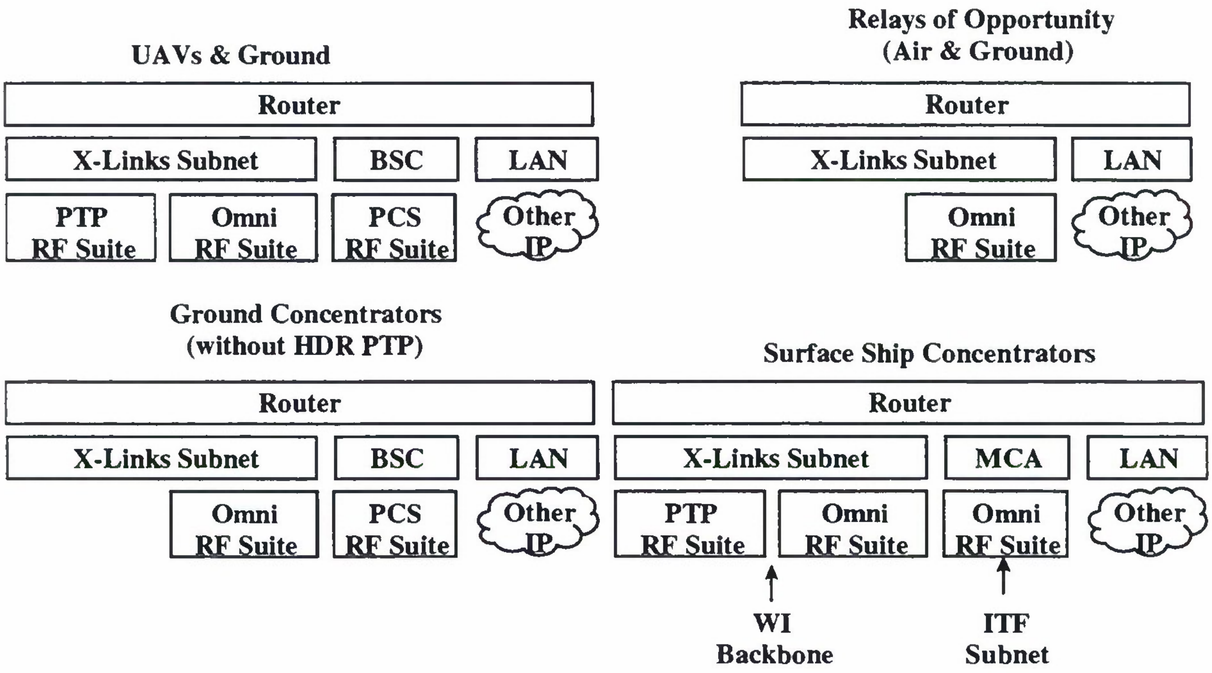


Figure 2. Platform configurations.

1.1 Cross-Link Subnet Concept of Operation

This section covers the concept of operation for the Cross-Links Subnet. We describe the network control structure and give examples of how data and voice traffic are handled.

The Cross-Links Subnet is a fully mobile wireless IP subnet, which means it handles its own internal routing. A key feature of the Cross-Links Subnet, which facilitates its ability to handle broadcast traffic, is the concept of a Network Structuring Algorithm (NSA) that dynamically develops and periodically maintains a backbone. This backbone has the feature that every node in the subnet is either on the backbone or is bidirectionally connected to a backbone node. Barring communication failures and assuming the topology of the subnet permits it, the backbone forms a connected, undirected graph. Figure 3 shows an example of a backbone that is formed within a multihop subnet. Nodes assume one of two roles—either they are backbone nodes (red) or non-backbone nodes (green or blue). Three types of links are shown: backbone links (red), backbone-connection links (blue), and ordinary links (green). As part of the reorganization process, every non-backbone node selects a particular backbone node to be its "backbone-connection node" (BCN). The link between a non-backbone node and its BCN is called a backbone-connection link. The BCN and the nodes that have selected it as their own BCN make up a Backbone-Connection Cluster (BCC).

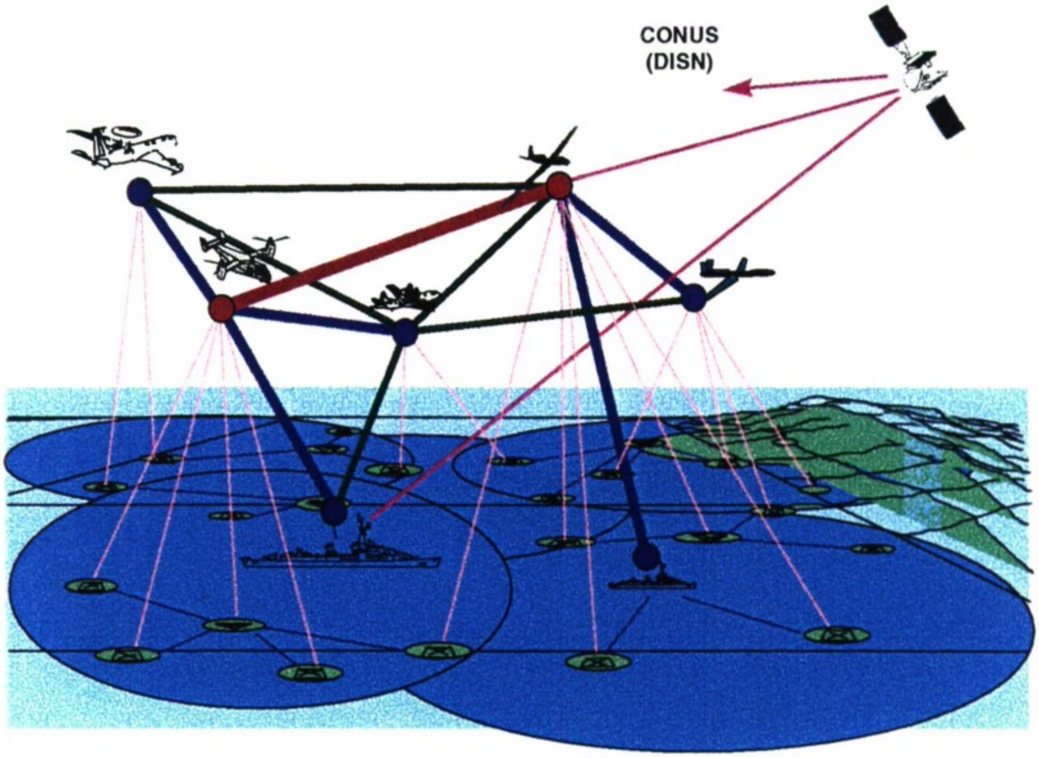


Figure 3. Example of the structure of a multihop Cross-Links Subnet.

The links of the backbone plus the backbone-connection links are of special note, because it is over these links that broadcast traffic is normally sent.

The Cross-Links Subnet can detect if the topology is fully-connected or star-connected. In the former, every node is bidirectionally connected to every other node; in the latter, the subnet is not fully connected, but there exists one or more nodes that are bidirectionally connected to every other node. In the case of a fully-connected subnet, no backbone is formed.

The network structure is somewhat different for star-connected topologies than it is for multihop topologies. Figure 4 shows an example of the backbone that forms in the case of a star-connected subnet. In that case, one or more (depending on the number of receivers available at each node) of the fully-connected nodes may become backbone nodes. Every non-backbone node is assigned one backbone node as its BCN. All the links between backbone nodes are considered to be part of the backbone also.

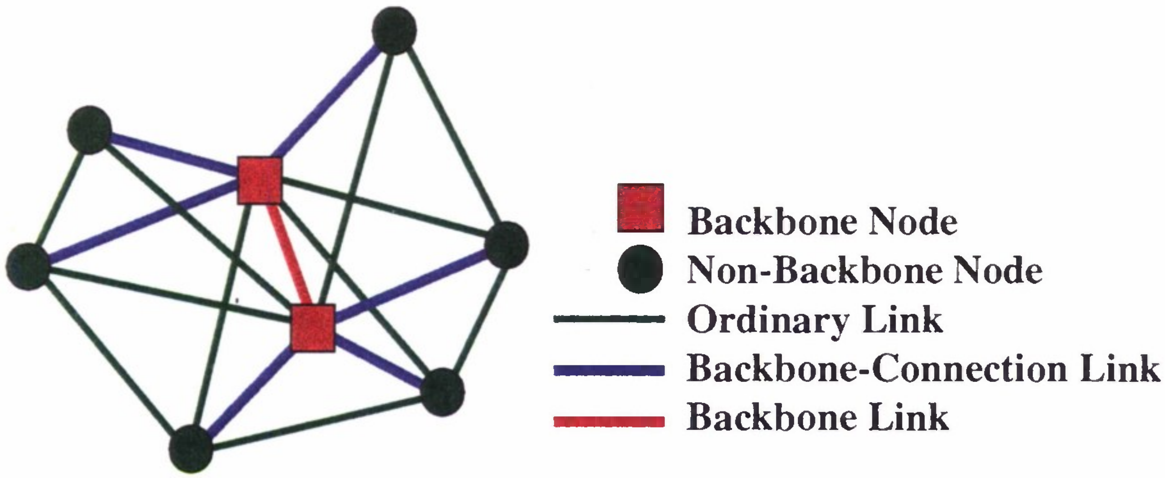


Figure 4. Example of the structure of a star-connected Cross-Links Subnet.

As mentioned earlier, the backbone and the BCN links are used to send broadcast traffic. Cross-Links Subnet nodes broadcast their local link types so that each node may have knowledge of the structure of the entire subnet. This database of link types is used to form two routing tables, which, in turn, are used for selecting the next relay for point-to-point traffic. One routing table is biased in favor of using backbone nodes, while the other is based on fewest hops to the destination. Congestion control determines the relative amount of traffic that is routed over the backbone versus how much is routed over the shortest (i.e., fewest hops) path. Point-to-point message headers contain a field to indicate which routing type (i.e., backbone or fewest hops) is being used for that message.

The Network Structuring Algorithm (NSA) uses the omnidirectional transmissions to identify reliable, bidirectional links. These links are then organized into backbone links, backbone-connection links, and ordinary links. Based on this classification of links and on traffic loads, the NSA sets up a set of high-data-rate links to effectively handle the network's traffic requirements. The NSA effects these links by issuing commands to point the directional antennas. If a sufficient number of directional antennas are available, the NSA sets up high-data-rate links along the backbone. Each Backbone-Connection Node is responsible for setting up the high-data-rate links for those nodes that have selected it as their own BCN. The BCN tries to identify a set of high-data-rate links within its BCC such that there exists a high-data-rate path between all nodes of its BCC. If there are not enough high-data-rate links to meet these objectives, the high-data-rate links are time-shared according to traffic loading requirements. If there are additional high-data-rate links that can be set up after the above objectives have been met, then these, also, are time-shared according to traffic loading requirements. Figure 5 shows an example of how the directed links might be deployed for our example.

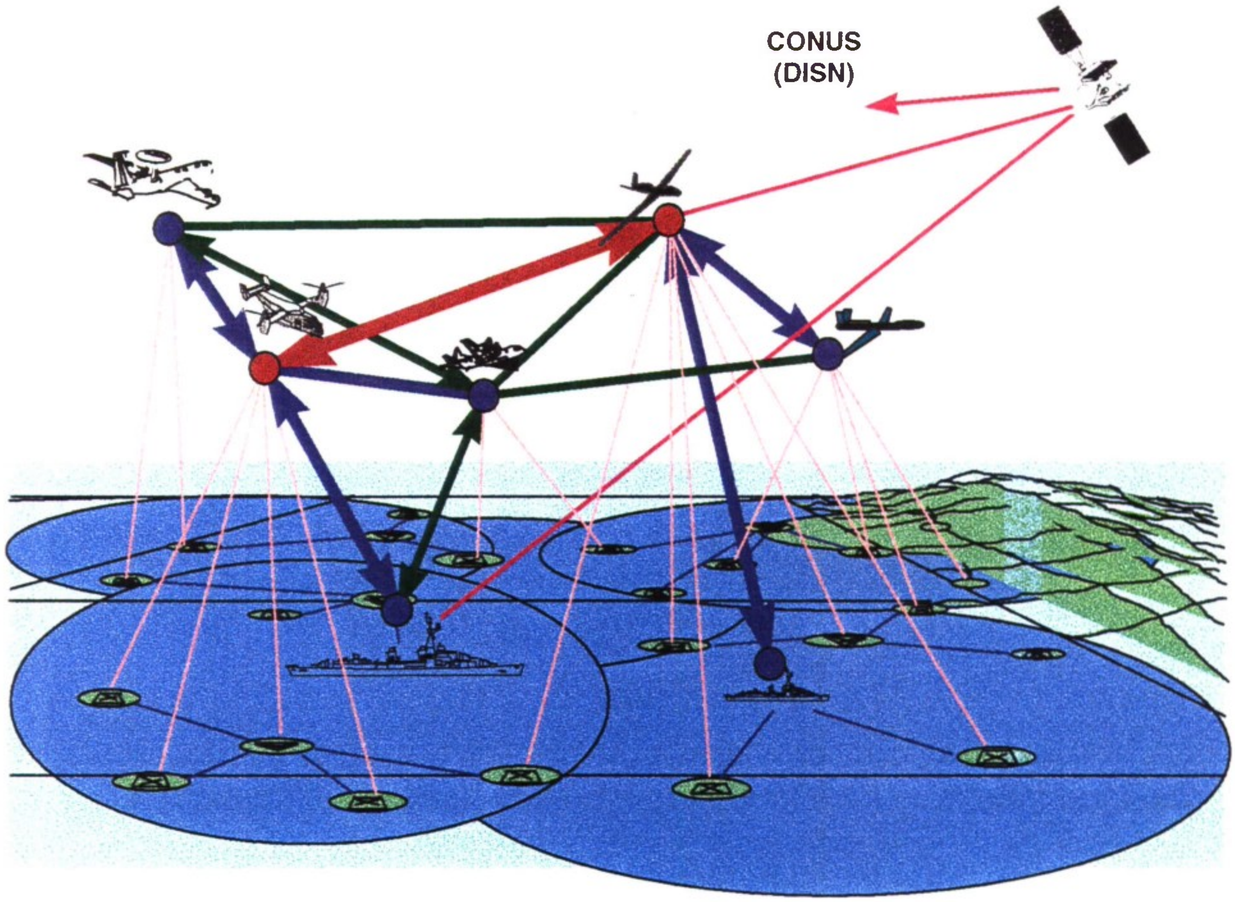


Figure 5. An example with directional links in place.

Traffic that is being routed over the backbone can be distributed to the non-backbone nodes belonging to a Backbone-Connection Cluster either via the BCN links or over the high-data-rate links set up within the BCC.

Figure 6 gives an example of the network structuring that takes place in the absence of one of the nodes of the Cross-Links Subnet.

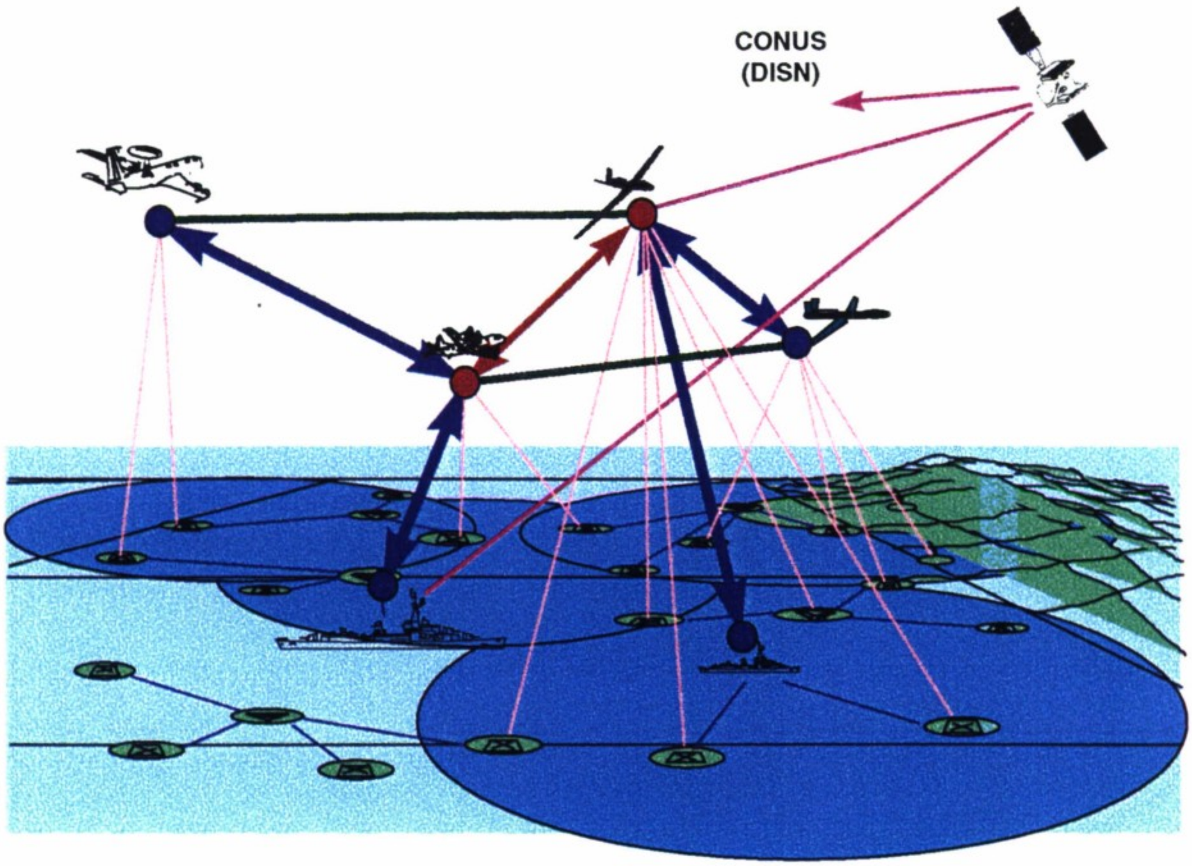


Figure 6. The example of Figure 5 minus one of the nodes of the Cross-Links Subnet.

The Cross-Links Subnet provides both virtual-circuit-switching and datagram-switching (i.e., packet-switching) capabilities. They are often referred to as connection and connectionless services, respectively. This dual capability is implemented as follows. The transmissions from Cross-Links Subnet transmitters are formatted into cells. Each cell is preceded by a cell header that includes a cell checksum and a field that indicates the type of cell. There are three types of cells, which are called synchronous, datagram-switched, and virtual-circuit-switched. The latter two types are sometimes lumped together under the classification of asynchronous cell. Virtual-circuit-switched cells are relayed to their destination prior to unpacking their contents, and they have transmission precedence over all but synchronous cells and the highest priority datagram-switched cells. Virtual-circuit-switching is especially useful to support interactive voice traffic and other real-time applications.

All services that are provided by the Cross-Links Subnet are set up by the Communication Services Manager (CSM) via the Subnet Provider Interface (SNPI); see Figure 7. The SNPI is located below the IP wireless interface and above the subnet. In order to set up a service, the service must first be registered by the Communication Services Manager. The registration process sets up two-way communication between the subnet "user" and the subnet, and it associates a (user-address, node-address) pair (called the Subnet Access Point or SNAP) with the user. Before using this service, the Communication Server must also specify whether the service is for datagram-switched or virtual-circuit-switched traffic.

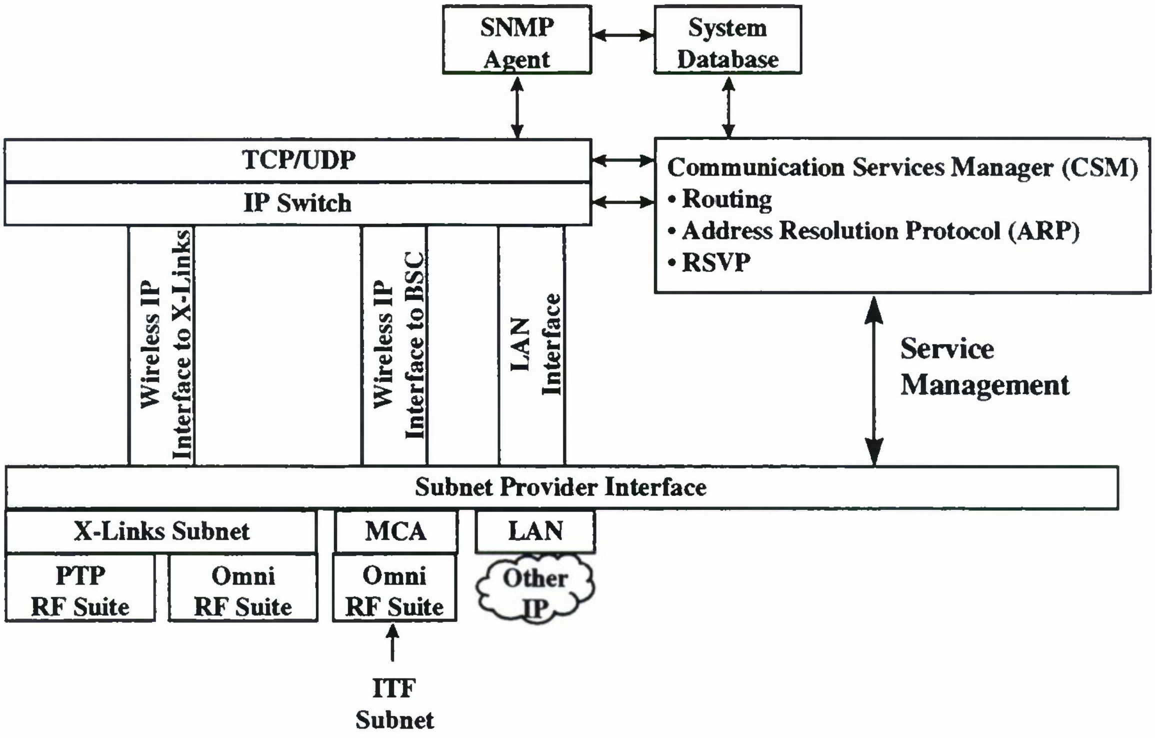


Figure 7. Surface ship concentrator.

As an example of the operation of the Cross-Links Subnet, we now consider how it receives and handles IP traffic that requires only best-effort delivery. Later we shall examine how the Cross-Links Subnet handles IP traffic that is associated with real-time delivery requirements, such as interactive voice. At startup time, the Communication Services Manager registers a service to handle all IP, best-effort-delivery traffic. In response to the registration request, the subnet sets up a Service Port for the requested service.

When the IP switch has a packet to send, it consults its routing table to determine to which interface to send this packet. If the packet is to be sent via the Cross-Links Subnet, the packet is sent to the wireless IP interface that is connected to this subnet. The wireless IP interface software must determine, from the source IP address, protocol number, and port number and the destination IP address and port number, which service handle to use to forward this packet to the SNPI and which SNAP to use for the destination. The wireless IP interface will consult the ARP (Address Resolution Protocol) database to obtain the node-address (i.e., the physical address) of the Cross-Links Subnet node that is the gateway for the destination. Armed with this information, the packet is forwarded to the SNPI. The SNPI then makes a call on the subnet to request that the subnet enqueue the IP packet.

When a packet is presented by the upper layer to the Cross-Links Subnet, the subnet first determines if it will accept the new packet. Flow control is handled by the Cross-Links Subnet using message priorities, message time-to-live, and associated "high-water" and "low-water" message queue levels. For each message priority level, there are associated high-water and low-water levels that control the number of new messages of that priority or higher priorities allowed at any time in the transmission queue. Once the Cross-Links Subnet node is fully operational, it will begin to accept traffic from the upper layer. When a packet of a given priority is submitted for transmission by the upper layer, the subnet checks to see if it is currently accepting packets of that priority level. If it is, the values of the variables representing the number of packets of equal or greater priority that are in the transmission queue are incremented by one. If this value reaches the high-water mark for a given priority level, additional new packets of that priority will not be accepted by the subnet until the corresponding low-water mark is reached. Whenever a new message is removed from the transmission queue (either because it was transmitted or its time-to-live was exceeded), the values of the variables representing the number of packets of equal or lessor priority that are currently in the transmission queue are decremented by one. When the low-water mark is reached, packets of that priority are accepted again.

If a packet is accepted by the subnet, an additional header is added to facilitate routing of this packet within the subnet. In particular, the user-address portion of the SNAP address of the destination node is prepended to the packet and a subnet message header is prepended to that. The subnet message header contains several fields including: message type, message length, message precedence, source node address, and message sequence number. Point-to-point message headers also include fields for the destination node, the relay node, and the routing type. If the node-address portion of the SNAP address is a unicast address, the packet is sent within a Cross-Links Subnet point-to-point message. If the node-address portion of the SNAP address is a multicast address, the packet is sent within a Cross-Links Subnet broadcast message. If the packet is being sent as a point-to-point message, the routing tables are consulted to determine the node-address of the relay node. Point-to-point datagram switched traffic can use either backbone or fewest-hops routes. Most relayed traffic is routed along the backbone. The decision to use backbone or fewest-hops routing is encoded into the routing-type field of the point-to-point message.

Once the packet has been accepted by the subnet and it has been incorporated into a Cross-Links Subnet message, it is put into the transmission queue associated with the SNAP address of the source. This queue is priority ordered, and within each priority, messages are serviced in FIFO (first-in-first-out) order. The transmission queue is part of the Service Port that is set up when the upper layer registers the best-effort-delivery service.

When a transmitter becomes idle, the Cross-Links Subnet software checks whether there is any traffic, in the form of a transmission cell in one of the service ports, waiting to be sent. If there is no synchronous cell nor any virtual-circuit traffic to be sent, a subnet looks for a cell from the service port associated with the best-effort-delivery service (i.e., the port holding messages that are using the datagram-switched service). A datagram-switched transmission cell can be filled once it is determined which transmitter will be used for the transmission and which nodes will be receiving this transmission. At that time, the highest priority message that is awaiting transmission to one of the receiving nodes is packed into the cell. If the cell is not filled, additional messages may be packed into the cell, space permitting. The message may also be transmitted over several cells.

Every node that receives a transmission, must unpack the Cross-Links Subnet messages from the stream of cells received. If the message received is a point-to-point message destined for another node, and if the receiving node is to serve as a relay of this message, then the next-hop destination field of this message is updated and the message is placed into the transmit queue. If the receiving node is the intended destination, the message payload is extracted and forwarded to the SNPI for delivery to the IP layer.

If the user application is real-time voice or some other real-time communication requirement, the Cross-Links Subnet will set up a virtual circuit to handle this service. First we consider the example where the user wishes to set up a point-to-point voice call. The user-application requests the service via the RSVP API. The resource needs are communicated from the RSVP demon to the Cross-Links Subnet via the SNPI interface. On receiving the resource request, the local Cross-Links Subnet node checks to see if the virtual circuit request can be handled locally. If it can, it performs several additional tasks required in preparation to setting up the circuit. The originating node then determines, from one of the point-to-point routing tables, the next node to the destination. The originating node sends a point-to-point message to the next-hop asking if it can support the required service. Each relay node checks to see if it can support the request. If it can support the request, the relay node will forward the request to the next relay node downstream. If it can't, it will send a fail indication upstream. When the destination node is reached, it checks to see if it can support the service. If it can, it sends a service-accept indication to the originating node. Assuming that the destination accepts the service, the destination node sends a message upstream that creates the circuit. When the source node receives the circuit-create message, it passes it to the Communication Services Manager. At this point the "user" can send traffic over the virtual circuit. It is possible for the application to request that a second "hot spare" circuit be established also. This spare is used in the event that the primary circuit is broken. In the event that the primary circuit is not broken, the full transmission capacity of the spare is available for other traffic. However, the use of "hot spares" does reduce the number of circuits that the network can handle at one time.

2. Cross-Links Subnet Capability Requirements

The Concept of Operation, just described, assumes a number of capabilities that the Cross-Links Subnet must provide. In this section, we describe these requirements.

Provide both datagram-switching and virtual-circuit-switching

- Implement a cell-based system in which cells are marked according to whether they are to be treated as datagram-switched or virtual-circuit-switched cells.
- Subnet messages packed in datagram-switched cells are unpacked on a hop-by-hop basis.
- Subnet messages packed in virtual-circuit-switched cells are only unpacked at the destination node(s).

Link quality monitoring

- Provide error detection capability at cell level.
- Monitor percentage of cells received correctly from each transmitter.
- Provide at least four levels for reporting link quality.
- Maintain a global database of link qualities for all RF systems.

Network self-organization

- Periodically identify all reliable, bidirectional links according to one of the following types: backbone link, backbone-connection link, and ordinary link.
- Periodically identify all nodes as one of the following types: backbone node or non-backbone node.
- The backbone shall have the following characteristics: each node shall be either a member of the backbone or bidirectionally coupled to a backbone node; the backbone nodes shall be connected by bidirectional links; the backbone shall be connected, if this is possible.
- Maintain antenna pointing for all directional antennas.
- Detect nodes joining and leaving subnet.
- Maintain a global topology database containing link types.

Virtual-circuit management

- Determine primary and (optionally) secondary paths for setting up point-to-point, virtual circuits.
- Upon request for establishment of virtual circuit, check whether network can support request.
- If network can support a new circuit, implement procedures to create and manage the circuit.
- Reclaim resources that have been reserved for virtual circuits, if these circuits have failed.
- Implement procedures to detect the failure of point-to-point, virtual circuits and effect procedures needed to switch to backup circuit (if such a backup service has been requested).
- Implement procedures to setup and maintain (in the presence of changing connectivities) broadcast and multicast services over virtual circuits.

Message packing/unpacking

- Pack/unpack user data into/from Cross-Links Subnet messages.

Cell packing/unpacking

- Pack/unpack subnet messages into/from cells.
- Must be able to handle combined high-data-rate and lower-data-rate links between neighboring nodes.

Cell multiplexing/demultiplexing

- User data awaiting transmission is stored in message queues within service ports. In addition, the subnet transmits control traffic at specific times in “synchronous” cells. When a transmission opportunity arises, the subnet first checks whether there is a synchronous cell ready to be sent. If there isn’t, the subnet queries one of the service ports for a cell.
- Cells must be selected in such a way that quality-of-service commitments are fulfilled.
- Multiplex/demultiplex cells onto/from several transmitters/receivers.

Flow and congestion control

- Implement priority-based flow control at SNPI interface.
- Balance traffic loads over multiple paths to avoid network congestion.

ARQ

- Provide ARQ service (selectable by user) that uses link-by-link ARQ at the cell level to achieve reliable subnet message transfer.

Manage message storage

- Provide a time-to-live for each message, and periodically remove those that have exceeded their time-to-live.
- Provide a pool of buffers to hold subnet messages.
- Detect and remove duplicate messages as they are received.

Provide routing

- Provide unicast and multicast routing for both datagram-switched and virtual-circuit-switched traffic.
- Provide primary and (optionally) secondary paths for unicast virtual-circuit services.
- Provide routing that can adapt to the loss of link quality; that is, the routing capability should work over links that may be one-way and/or unreliable.

Provide receiver control

- Each receiver must be tuned to the appropriate signalling channel, i.e., frequency-hop (FH) code.
- If a receiver is not required to monitor a specific channel, then it must be deactivated or tuned to a preassigned "NULL" channel.
- Compute and disseminate receiver schedules if receivers must be shared.
- Detect and avoid use of failed receive string.

Provide transmitter control

- This provides a capability to control when the transmitter is allowed to transmit. For the Cross-Links Subnet, the channel access protocol is trivial—a node is allowed to transmit continuously.
- Provide information on what nodes can be reached during a particular transmission and from a particular transmitter.

Provide antenna control

- There must be an API for controlling the high-data-rate directional antennas—both transmit and receive.
- This interface should permit an antenna to be directed by specifying the GPS coordinates of both the source and destination nodes.
- It is assumed that the antenna systems themselves will have access to any platform orientation information (i.e., yaw, roll, and pitch) that is necessary to carry out the prescribed commands.
- Provide capability for manual override of automated antenna pointing.

Provide modem control

- Implement modem API that allows the subnet to change the data rate over a link in response to changing link conditions.

Provide link encryption

- Link encryption at rates that lie in the range from 100s of kbps to several Mbps are needed.
- Automatic resynchronization of cryptos is required over links that may be unidirectional or bidirectional.

Handle SNPI interface

- This capabilitity enables the subnet to communicate over the SNPI interface to the Communication Server. The subnet coordinates with the CSM to set up the various communication services requested. After a service is set up, the subnet takes traffic from the SNPI interface and places it in the appropriate service ports for subsequent transmission.
- The subnet, when it receives traffic destined for the upper protocol layer at this node, must deliver this traffic via the SNPI interface.

APPENDIX B

PROTOCOLS TO SUPPORT INTEGRATED SERVICES IN A MOBILE CELLULAR SYSTEM ARCHITECTURE (FOR ARCHITECTURE OPTION 2)

1. Introduction

Cellular systems are designed to provide a fixed amount of communication bandwidth to each mobile subscriber. Media access in digital systems is managed either by CDMA or a combination of CDMA and fixed TDMA. This works well when the system is servicing voice users that require uniform bandwidth reservations. However, this approach becomes inefficient when users are sending bursty data traffic. Cellular providers are developing at least two sets of standards that address this issue—General Packet Radio Service (GPRS, European) and Cellular Digital Packet Data (CDPD, US). Our approach differs significantly from both of these emerging standards. We propose a single, integrated cellular system that accommodates both data (i.e., bursty, non-real-time datagram) and voice (i.e., real-time, virtual circuit) services with the ability to both reserve and dynamically reallocate bandwidth according to a user's current usage. We leverage previous work done at NRL in support of the Data and Voice Integration ATD. That work developed a unique cell multiplexing scheme that creates virtual channels for both datagram and virtual circuit services. We adapt this scheme to a cellular architecture by designing a new Medium Access Layer protocol for mobile subscriber transmissions.

The goal of WI is to create a seamless, highly mobile, communication infrastructure that can be rapidly deployed and automatically configured during the early phases of a combat operation. Likely operational scenarios include littoral operations, small unit operations, rapid deployment of tactical forces, and special operations forces. Desired communication capabilities include a full spectrum of integrated services from real-time voice and video services, pseudo real-time services such as telnet or interactive whiteboard, and non-real-time services, e.g., web browsing, FTP, email, etc.

The current consensus on WI communication architecture is:

- WI applications will be IP-based.
- The WI architecture will necessarily contain IP routers.
- WI will use airborne relays to provide BLOS connectivity to its subscribers.
- WI relays will be interconnected by medium data rate (~100s kbps) omnidirectional links as well as high data rate (~10s Mbps) directional links.
- Some WI relays may be ground-based (i.e., ground concentrator nodes).
- WI mobile subscribers use specially developed Mobile Communications Devices (MCDs) for acquiring WI communication services.

- Interconnection with ATM switches is a desired capability.

Details of the WI architecture presented in this paper represent NRL's attempt to construct a comprehensive and consistent design that meets all the WI design goals. NRL's contributions to the WI architecture are particularly strong in the areas of managing mobility and supporting integrated services [1]–[3]. Our focus is upon those portions of the WI architecture that provide the ability to support integrated services. The remainder of this appendix describes a high level design and concept of operation of protocols to support integrated services in a mobile cellular system architecture (for architecture option 2), and then presents more detailed designs of the mechanisms used to create Quality of Service (QoS) capable subnetworks.

2. Warfighter's Internet Architecture and Concept of Operation

The Warfighter's Internet architecture consists of three major system segments:

- IP Router
- Cross-Links Subnet
- Cellular/PCS Subnet

Figure 1 depicts a littoral scenario, showing a WI topology constructed from links of both types, i.e., Cross-Links Subnet links and Cellular/PCS Subnet links. The Cross-Links Subnet creates a high capacity, highly mobile backbone for the WI system, using both high capacity directional links and medium capacity omnidirectional links. As one might suspect, the omnidirectional links are used to configure the directional links. However, what is not obvious is that the omnidirectional subsystem is fully capable of supporting a full range of integrated communication services just like the directional subsystem is, only with less communication bandwidth. Thus, the Cross-Links Subnet uses both subsystems in parallel, combining the capacity of both to transport WI backbone traffic. As is shown in Figure 1, WI backbone nodes, i.e., those interconnected via the Cross-Links Subnet, are located on some surface platforms as well as on airborne platforms.

Most (but not necessarily all) WI backbone nodes are collocated with Cellular/PCS Subnet Mobile Base Stations. The Cellular/PCS Subnet is comprised of two types of nodes: 1) Mobile Base Station (MBS) nodes and 2) Mobile Subscriber (MS) nodes. MS nodes do not communicate directly with other MS nodes; rather, all communication in the Cellular/PCS Subnet is relayed by MBS nodes that create a star topology with their locally attached MS nodes. Each active MS node is affiliated with one, and only one, MBS node at a given point in time. As MBS and MS nodes move relative to one another, the affiliations between MBS and MS nodes dynamically change. Figure 2 shows Cellular/PCS Subnet system components.

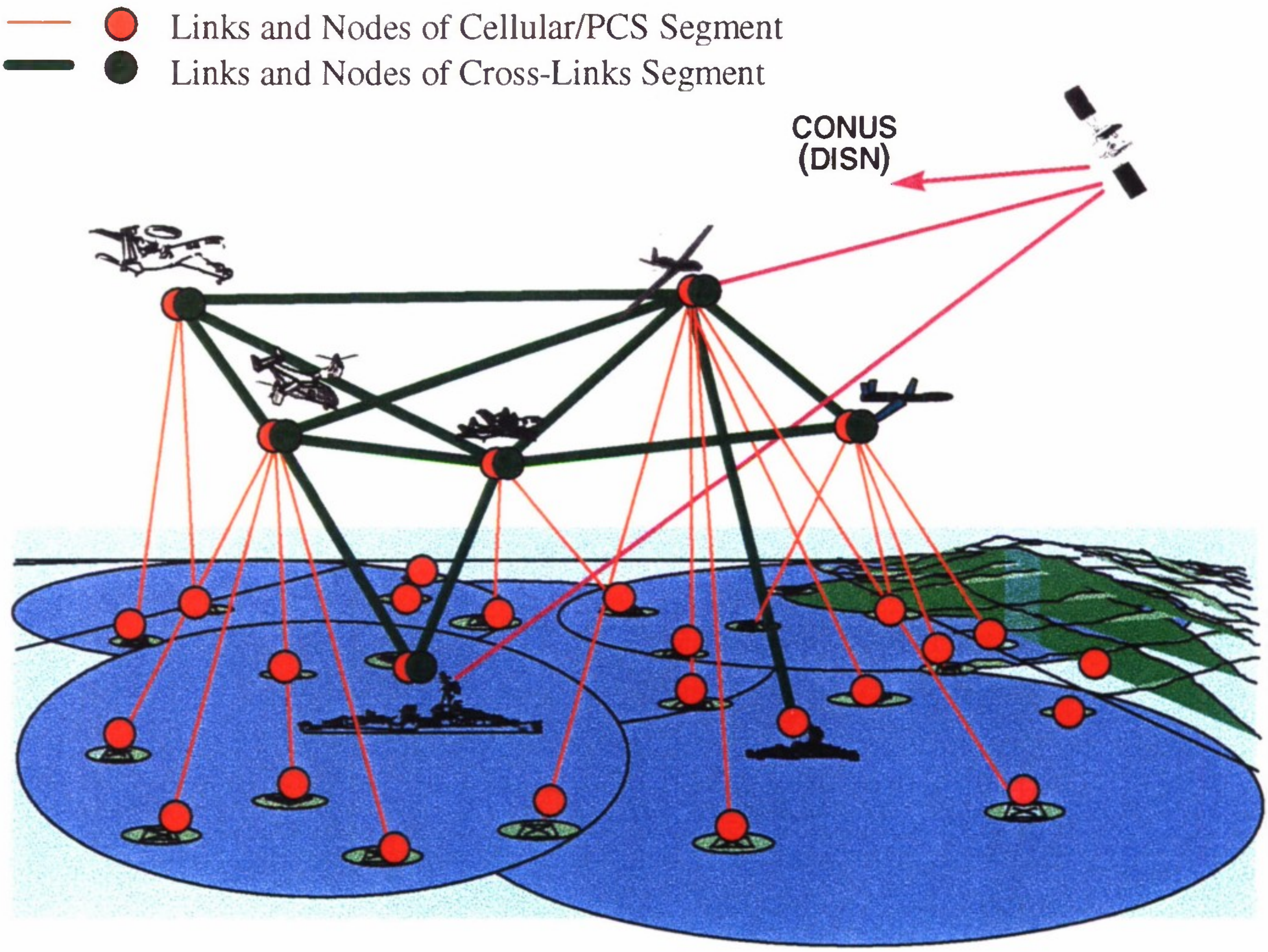

Links and Nodes of Cellular/PCS Segment
Links and Nodes of Cross-Links Segment
CONUS
(DISN)

Figure 1. WI topology for a littoral scenario.

An MBS node maintains full duplex connectivity with its affiliated MS nodes by using a high data rate downlink that broadcasts to all MS nodes in the footprint of its downlink modulator beam and by supporting a suite of uplink demodulators that receive the uplink transmissions of all affiliated MS nodes within the downlink footprint. The coverage area of the downlink beam's footprint is called a "cell." A single MBS may support communication in multiple cells by managing multiple downlink modulators and multiple suites of uplink demodulators. Management of the modulator/demodulator hardware used to service a single cell is facilitated by a Base Station System (BSS). Thus, one MBS may have several associated BSS components, each of which handles full duplex communication within its cell. MS nodes, therefore, are affiliated at a point in time with one, and only one, cell. When an MS node changes its cell affiliation, a "handover" is said to have occurred. Two levels of affiliation exist: 1) the MS/cell affiliation and 2) the MS/MBS affiliation.

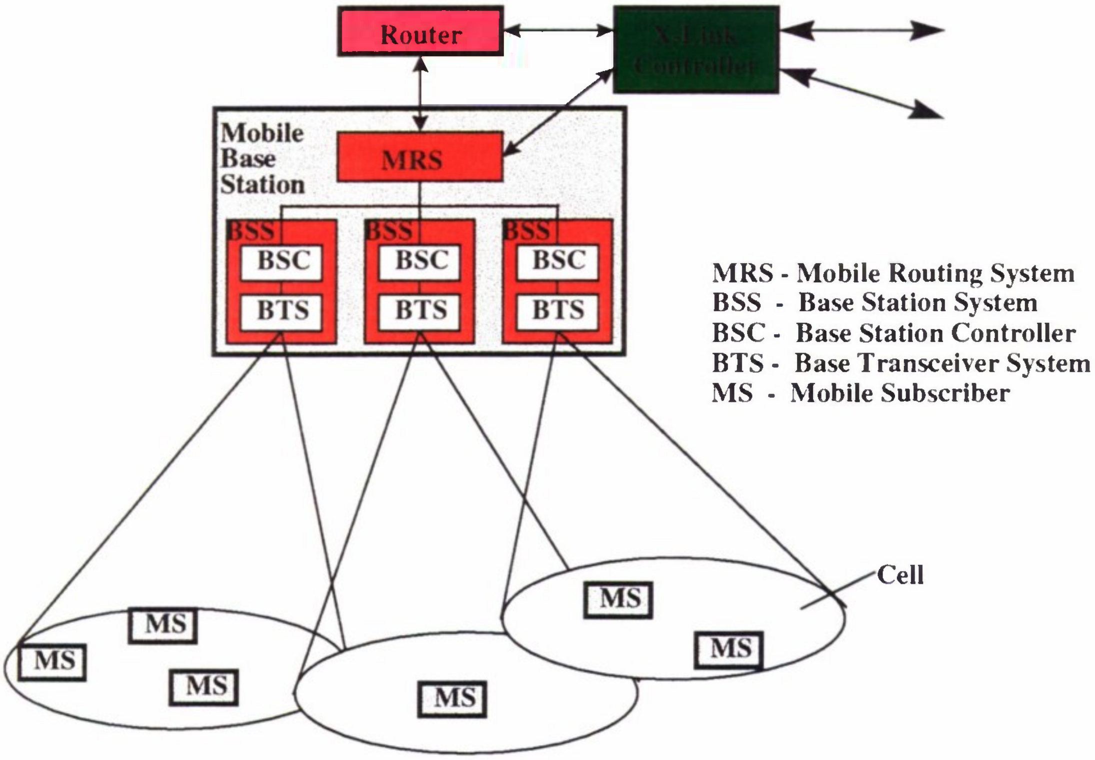

Router
X-Link Controller
Mobile Base Station
MRS
BSS
BSC
BTS
BSS
BSC
BTS
BSS
BSC
BTS
MRS - Mobile Routing System
BSS - Base Station System
BSC - Base Station Controller
BTS - Base Transceiver System
MS - Mobile Subscriber
Cell
MS
MS
MS
MS
MS
MS

Figure 2. Subsystems in the Cellular/PCS Subnet architecture.

It is possible for the number of MS nodes within a cell to exceed the number of demodulators in an uplink demodulator suite. Therefore, subscriber uplinks must be capable of timesharing the uplink demodulators. This capability is provided by a unique type of demand assigned multiple access protocol called Shared Resources Multiple Access (SRMA). SRMA dynamically adapts to the instantaneous bandwidth needs of each MS node based on both the offered traffic load at the MS node and QoS commitments made by the MS node.

Full connectivity within the Cellular/PCS Subnet is obtained by interconnecting all the MBS nodes. This interconnection is made possible by the Cross-Links Subnet. A full discussion of the Cross-Links Subnet is in Appendix A. Suffice it to say that the Cross-Links Subnet hides its rapidly changing topology from its users. To its users it appears to be a static subnet. The Cross-Links Subnet provides subnet transport services to an IP router, possibly to an ATM switch, and to the MBS. The MBS uses the cross-links transport services to send both Cellular/PCS Subnet signalling and data packets. At a minimum, signalling is used to maintain a distributed database which describes all current MBS/MS affiliations. This database is used by the MBS to make routing decisions for each data packet. Data packets that must be routed to a remote MBS for delivery to an MS affiliated with the remote MBS are transported via the Cross-Links Subnet.

The Cellular/PCS Subnet is a packet-switched network that supports both datagram and virtual circuit services. These services are obtained from the Cellular/PCS Subnet via the Subnet Provider Interface (SNPI). This is the same interface and service model supported by the Cross-Links Subnet. Since the Cellular/PCS Subnet uses the Cross-Links Subnet for data transport among the MBS nodes, the Cellular/PCS Subnet cannot make QoS commitments beyond what the Cross-Links Subnet can also support. The Cellular/PCS Subnet uses admission control, flow control, the QoS support of SRMA, cell multiplexing and adaptation layer protocols (i.e., ISCM—Integrated Services Cell Multiplexing), and the QoS capability of the Cross-Links Subnet to provide the services offered at the SNPI.

The Cellular/PCS Subnet is a fully mobile, wireless subnet that handles its own routing. As such, it is able to hide the mobility of all MS nodes from the IP Layer. To the IP Layer, the Cellular/PCS Subnet appears to be a static IP subnet which has the added benefit of being QoS capable. IP addresses can be statically assigned to each IP interface connected to a Cellular/PCS node of either type—MS or MBS. The mapping between IP and cellular addresses can be dynamically obtained via ARP (Address Resolution Protocol) or, conceivably, could be downloaded as part of a node's Complan.

3. Subnet Provider Interface (SNPI)

The Cross-Links Subnet and the Cellular/PCS Subnet both support the same service model and are accessed by the same interface. This common interface is called the Subnet Provider Interface (SNPI). Figure 3 shows that this interface is located directly underneath IP in the protocol stack. Since the SNPI is a subnet interface that is QoS capable, it is non-standard. Thus, in Figure 4, RSVP, RARP, ARP, and IP's lower interface are shaded differently to indicate that some modification was necessary to interface with the SNPI.

The design decision that motivated the SNPI's development was the decision to build QoS capability directly into the Subnet Layer as opposed to attempting to create QoS capability at a higher layer. In the Resource Reservation Protocol (RSVP) Functional Specification, the Packet Scheduler and Admission Control functions are shown residing above the Subnet Layer (Figure 4). In order for the Packet Scheduler to provide QoS capability to the data flows identified by the Classifier function, the Packet Scheduler must interface to a subnet that provides either 1) a guaranteed bandwidth, 2) is over-provisioned, or 3) has built-in QoS capabilities. In the third case, the Packet Scheduler is redundant, as is the Admission Control function. Presuming either the first or the second case in a mobile, wireless system is very optimistic. Thus, for the Warfighter's Internet, building QoS capability directly into the Subnet Layer is the option that makes the most sense. The SNPI is a well-defined interface for accessing the services provided by a QoS capable subnet.

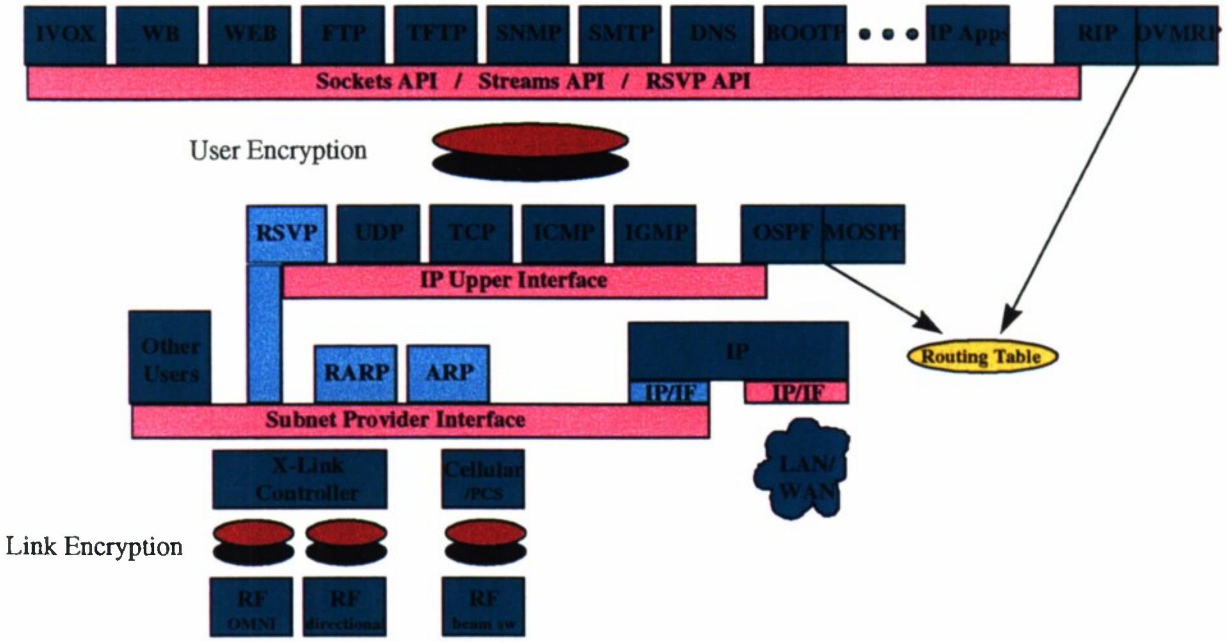


Figure 3. Location of the Subnet Provider Interface in the WI protocol stack.

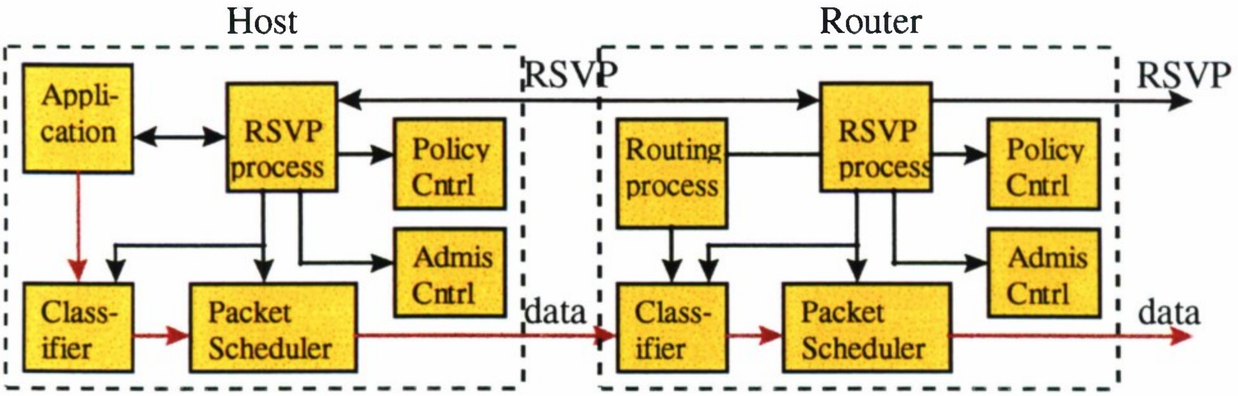


Figure 4. RSVP hosts and routers (Internet Draft—RSVP Functional Spec.).

The SNPI supports the following functions:

- **User Registration** {The registration process identifies each user to the Subnet Layer. IP, ARP, RARP, and other users each have a unique ID. This ID enables the Subnet Layer to identify and deliver traffic to multiple users. *One significant point not directly illustrated in Figure 4 is that the Cellular/PCS Subnet's MRS component is a user of the Cross-Links Subnet's SNPI and vise-versa.* This is shown in Figure 2 by the line that interconnects the X-Link Controller and the MRS.}

- **Service Setup and Termination**

  – *QoS Negotiation* {The SNPI supports a QoS negotiation strategy. If a user request for a subnet service fails (i.e., because of admission control), the subnet can include acceptable QoS parameters in the service denial response. The user then can decide whether or not to issue a new service request with acceptable QoS parameters.}

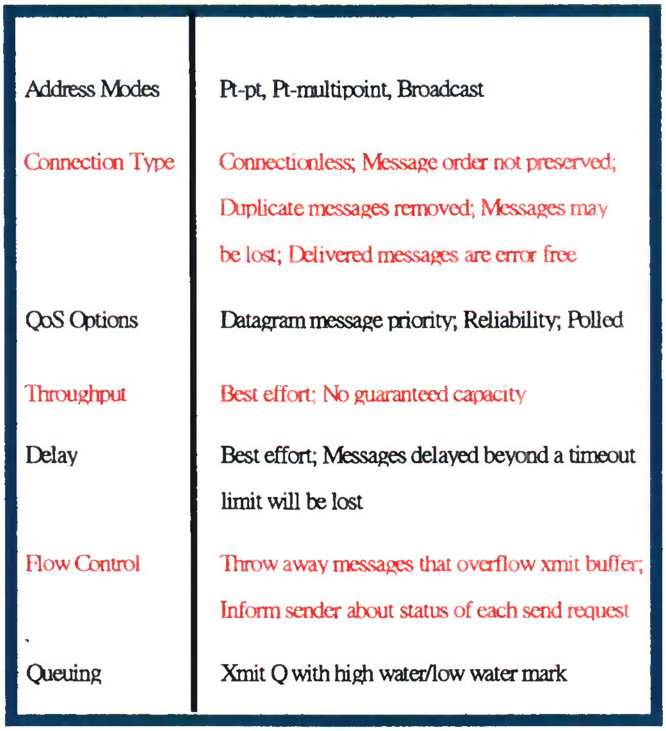
  – *Admission Control* {The Subnet Layer must first determine if the resources are available in order to honor a request for a service that reserves Subnet Layer resources (i.e., virtual circuit service request).}

  – The service setup and termination capabilities provided by the SNPI can facilitate service preemption by higher layer protocol where preemption rules can be determined. These rules should not be built into the Subnet Layer.

- **Subnet Layer Data Transport** {The SNPI supports two broad categories of data transport service—datagram and virtual circuit. Both types of service are flow controlled.}

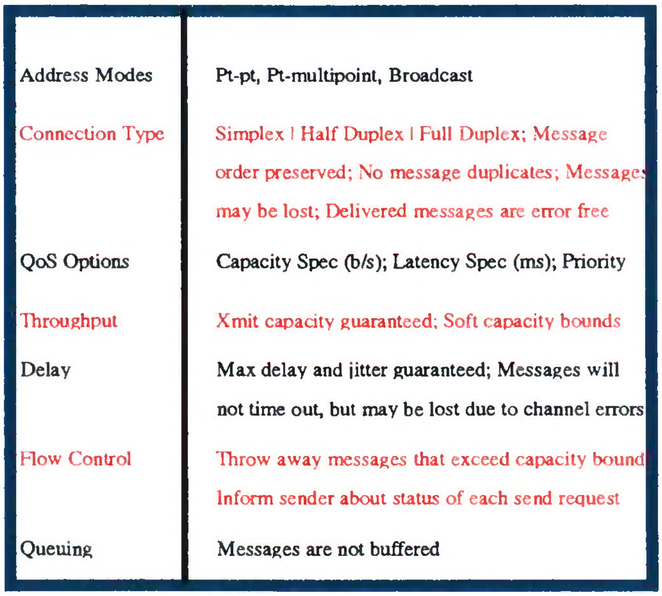
- **Subnet Layer Status** {The subnet can accept status queries as well as report a large quantity and variety of status information via the SNPI.}

Figures 5 and 6 briefly describe the service models for the two major categories of service provided via the SNPI, i.e., datagram service and virtual circuit service.



Address Modes	Pt-pt, Pt-multipoint, Broadcast
Connection Type	Connectionless; Message order not preserved; Duplicate messages removed; Messages may be lost; Delivered messages are error free
QoS Options	Datagram message priority; Reliability; Polled
Throughput	Best effort; No guaranteed capacity
Delay	Best effort; Messages delayed beyond a timeout limit will be lost
Flow Control	Throw away messages that overflow xmit buffer; Inform sender about status of each send request
Queuing	Xmit Q with high water/low water mark

Figure 5. Datagram service model.



Address Modes	Pt-pt, Pt-multipoint, Broadcast
Connection Type	Simplex | Half Duplex | Full Duplex; Message order preserved; No message duplicates; Messages may be lost; Delivered messages are error free
QoS Options	Capacity Spec (b/s); Latency Spec (ms); Priority
Throughput	Xmit capacity guaranteed; Soft capacity bounds
Delay	Max delay and jitter guaranteed; Messages will not time out, but may be lost due to channel errors
Flow Control	Throw away messages that exceed capacity bound Inform sender about status of each send request
Queuing	Messages are not buffered

Figure 6. Virtual circuit service model.

4. Integrated Services Cell Multiplexing (ISCM)

ISCM defines a Cell Multiplexing Layer and an Adaptation Layer as the mechanisms used to support integrated services (IS) and to provide the QoS capabilities for the Cellular/PCS Subnet. In the Cellular/PCS Subnet, multiplexed flows of cells are transmitted on both the uplinks and the downlinks. Cells are classified as datagram (DG) cells or as virtual circuit (VC) cells. Virtual circuit cells are further classified as data or as signalling cells. VC signalling cells create virtual channels on both uplinks and downlinks for setting up and tearing down virtual circuits on those same links. VC data cells are further differentiated by a VC number. The net effect is that virtual channels are created on both uplinks and downlinks for transporting datagram traffic (one channel per link), virtual circuit traffic (multiple channels per link), and VC signalling (one channel per link). This concept is illustrated in Figure 7.

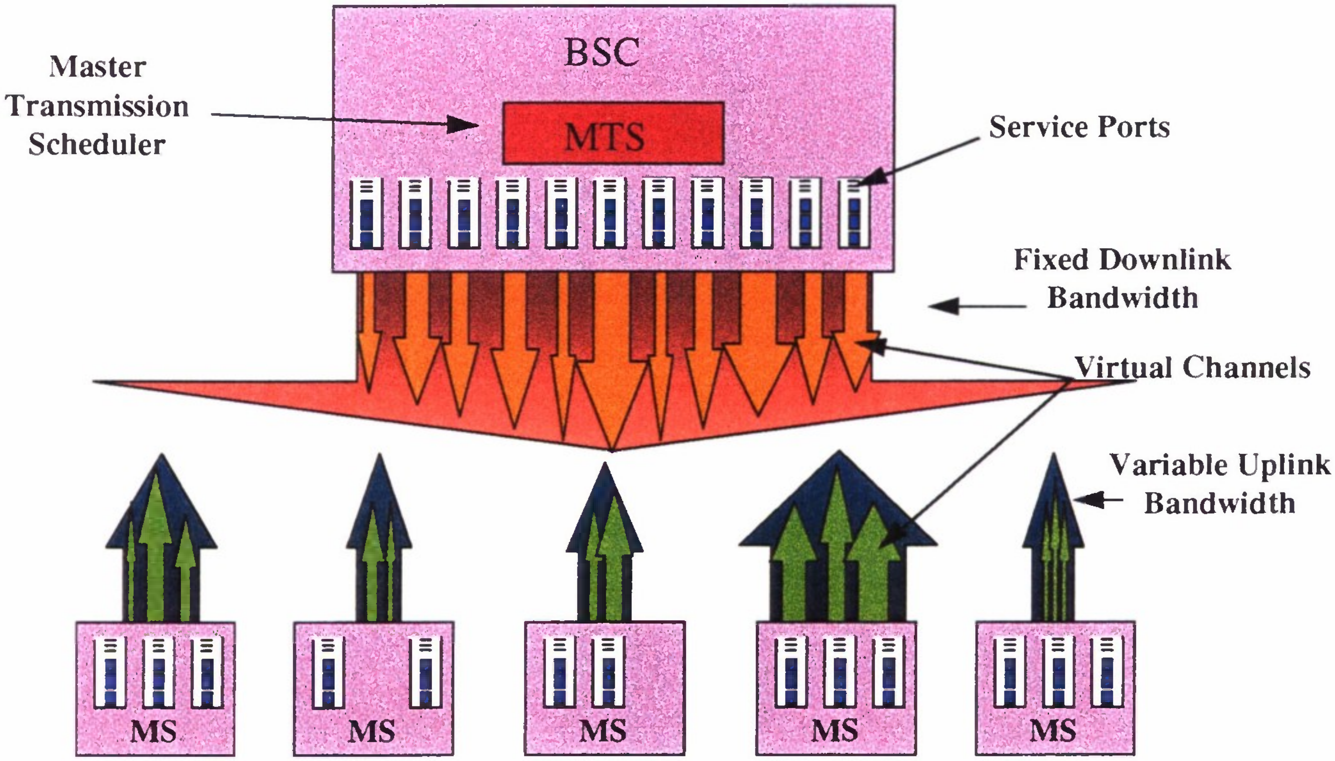


Figure 7. Uplink and downlink virtual channels created by cell multiplexing.

In many respects, this is similar to the ATM (Asynchronous Transfer Mode) approach— i.e., using cells to create virtual channels which, in turn, support QoS commitments. One difference between ISCM and ATM is in the way the Adaptation Layer messages are packed into cells. In the ATM approach an Adaptation Layer message is always packed into an integral number of ATM cells. In most cases this requires padding to fill out the cell that contains the message trailer. For a long Adaptation Layer message, the Adaptation Layer overhead ([trailer + padding] / length) will be small. However, when Adaptation Layer messages are short, Adaptation Layer overhead can approach 50%. By contrast, in ISCM, Adaptation Layer message packing is completely independent of cell boundaries. A message header (ISCM uses message headers rather than message trailers) can begin anywhere in a cell payload. In fact, the header itself can be split across a cell boundary. Partially filled cells are padded and flushed automatically as excess bandwidth becomes available; or an SNPI user may call a flush command explicitly. This approach helps reduce overhead and conserve bandwidth.

In ISCM, each virtual channel and, therefore, each distinguishable cell type is associated with its own Service Port. This feature is also shown in Figure 7. The service ports themselves are serviced by the Multiplexer. The cell multiplexing control loop operates on the service ports that have been created and placed in the Service Port Queue (see Figure 8). One service port is created to handle all datagrams, one service port is created to handle all Subnet Controller peer-to-peer commands (i.e., VC signalling), and service ports are dynamically created to handle virtual circuits.

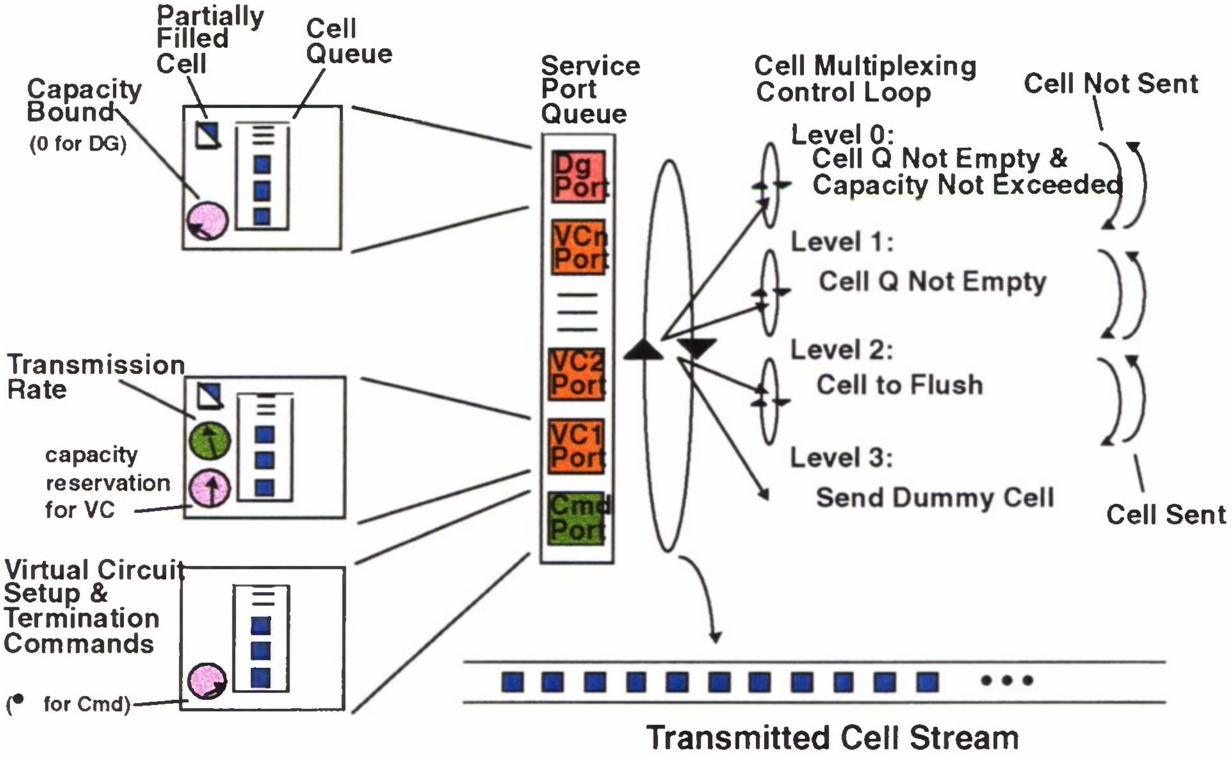


Figure 8. Multiplexing cells to create a transmitted cell stream.

Every service port has a FIFO cell queue and a capacity bound, i.e., a limit that is used to control the transmission capacity allocated to the service port. The DG Service Port's capacity bound is always set to zero, meaning that no capacity is reserved for datagram service. The VC Signalling Service Port's capacity bound is set to infinity, meaning that Subnet Controller peer-to-peer commands will get all the capacity that the transmission system can provide, if needed. In actuality, Subnet Controller commands use only a very small fraction of the transmission capacity. VC Service Ports have capacity bounds that are set in accordance with user requests for capacity that have been honored by the Subnet Controller. In addition, the DG and VC Service Ports may each have a partially filled cell that is in the process of being packed and has not yet been moved to the cell queue. It is moved to the cell queue when packing is completed or when the service port's flush() function is called. Also, the VC Service Port measures the user's transmission rate.

The service port entities just described are used by the cell multiplexing control loop to select the next cell for transmission and move it to the transmitted cell stream. The loop continually traverses the linked list of service ports in the Service Port Queue in an attempt to find a cell that meets the criterion for transmission. The search begins at Level 0, which uses the most restrictive test to determine if a cell can be sent. At Level 0, the cell queue must have a cell available to send and the capacity bound must not be exceeded if the cell is transmitted. The entire linked list of service ports is traversed using the Level 0 test. One cell is transmitted in turn from each service port that meets the Level 0 test criterion. If no cell is found that can be sent at Level 0, the level is incremented and the Service Port Queue is again traversed. At Level 1, the test criterion is relaxed so that a service port may send a cell if its cell queue contains a cell regardless of the port's capacity bound. If the entire Service Port Queue is traversed at Level 1 without finding a cell to send, the level is again incremented. At Level 2, partially filled cells are flushed and at Level 3 a dummy cell is sent. (In the Cellular/PCS Subnet, it is inappropriate to transmit dummy cells. Dummy cell transmission is appropriate when the link is synchronous and must send something even if no data is available for transmission.) If traversal of the list at a given level results in the transmission of at least one cell, the level is decreased by 1.

The end result of the multiplexing protocol just described is that cells are removed and transmitted from eligible service ports in a round-robin fashion. The Signalling Service Port and VC Service Ports that are not exceeding their capacity reservations are serviced first. The DG Service Port can only be serviced at Level 1 or higher, i.e., after the Signalling and VC ports have received their allocated capacity, if needed. If VC ports don't use all the capacity they have been granted, then that unused capacity is automatically transferred to support the DG Service Port. Also, automatic flushing is performed as capacity permits.

5. Mobile Routing System (MRS)

The MRS is one of the two major subsystems that comprise an MBS (Mobile Base Station—see Figure 2). In the "ping" examples given in Section 6 of the Appendix, many of the MBS actions described are performed by the MRS subsystem. Figure 9 gives a more detailed view of the MRS. We are not providing a complete description of all the transactions, e.g., handovers needed for the MRS function. We intend to use the applicable cellular technology available from COTS packet switched Cellular/PCS systems such as GPRS or CDPD when they become available.

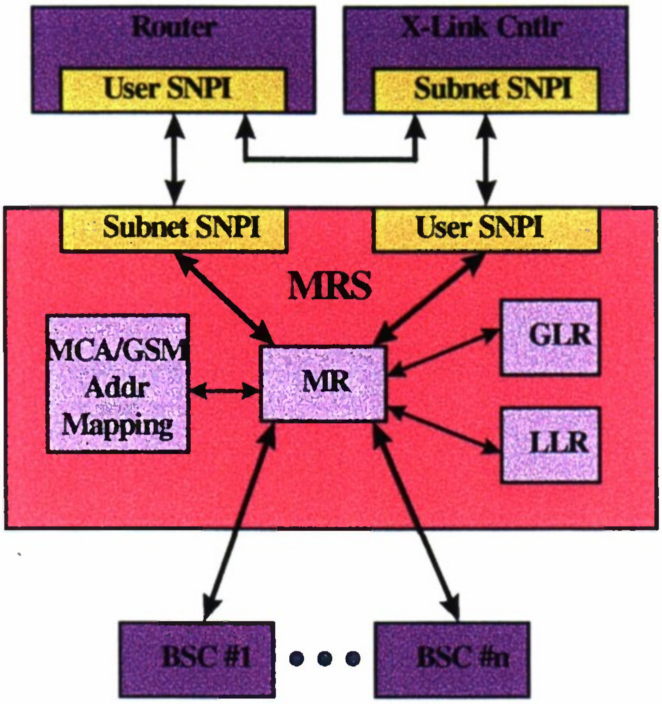


Figure 9. Mobile Routing System (MRS) components.

5.1 MRS Interfaces

The MRS has an interface to each of the other WI backbone node systems, i.e., the Cross-Links Controller and the IP Router. Also, the MRS has an interface to each Base Station Controller (BSC), which are MBS subsystems.

5.1.1 Interface to Router

The MRS interface to the Router is the *subnet* part of the SNPI described in section 3. This is the interface that the Router uses to access the QoS capable data transport services offered by the Cellular/PCS Subnet. Also, the MRS uses this interface to deliver traffic to the Router.

5.1.2 Interface to Cross-Links Controller

The MRS interface to the Cross-Links Controller is the *user* part of the SNPI. The MRS uses the interface to access the QoS capable data transport services offered by the Cross-Links Subnet. The MRS needs the Cross-Links Subnet services to transport both signalling and data to peer MRS entities at remote WI backbone nodes.

5.1.3 Interface to BSC (data)

The MRS' Mobile Router component interfaces to each BSC component of the same MRS. This interface handles data flow. The data flow from the BSC to the MR is traffic collected from all the Mobile Subscriber (MS) uplinks managed by the BSC. The uplink flow at the point it traverses this interface is composed of Cellular/PCS packets that are carrying IP packets as payloads. The flow will also carry VC cells since the Cellular/PCS Subnet will transport VC cells all the way to their destination address before unpacking. Downlink traffic traverses the interface from the MR to the BCS. This data traffic consists of Cellular/PCS packets (datagrams) and VC cells, since VC cells are transported end-to-end.

5.1.4 Interface to BSC (signalling)

If the MRS performs User Authentication and Equipment Identification functions for the BSCs, some signalling between the MRS and the BSCs will occur.

5.2 MRS Databases

Figure 9 calls out three MRS databases: the Global Location Register, the Local Location Register, and MCA/GSM address mapping. Other databases are implied by the User Authentication and Equipment Identification functions previously mentioned.

5.2.1 Global Location Register (GLR)

Each MRS maintains a GLR database that indicates for every MS which MBS it is directly connected to. All GLR databases are simultaneously updated by using the Cross-Links Subnet broadcast capability. Updates can be triggered by handover events that result in changes in MS/MBS affiliations. To limit the number of updates being broadcast, an MS must be actively receiving data when the handover occurs to trigger the broadcast. Otherwise, updating is triggered by an MBS when it receives a misdirected packet.

5.2.2 Local Location Register (LLR)

The MRS maintains a LLR database in order to switch downlink traffic to the correct BSC. The LLR identifies for each locally connected MS which BSC is servicing the connection. This database is updated whenever a local handover occurs (i.e., a handover involving an MS that is locally connected or was locally connected before the handover occurred).

5.2.3 MCA/GSM Address Map

For each MBS type of node in the Cellular/PCS Subnet, this database indicates which Cross-Links Controller it is directly interfaced to. This is a static mapping.

5.3 Mobile Router (MR)

The MR performs all the switching functions within the MRS. It uses the information in the databases just described and the Cellular/PCS destination addresses of the traffic it is switching (datagram packets or VC cells) to direct the traffic to the proper destination.

6. Examples

In these examples we describe the mobile communication system concept of operation by showing how a "ping" would work. The ping program sends an ICMP echo request message to a destination host, expecting an ICMP echo reply to be returned. It tests IP connectivity to the destination host. All the mobile subscribers are shown to be in the same Class B subnetwork. Even though the example shows the Mobile Communication Device (MCD) as an IP capable device, it is not mandatory (the MCD could be a Freewave type of radio connected to the notebook computer where the IP capability resides). The example shows the Mobile communication system using a Global Location Register (GLR), which maintains an accurate Mobile Subscriber (MS) to Mobile Base Station (MBS) mapping. We assume that the MBS is attached to only one BSS and, therefore, we can simplify the example by ignoring the Local Location Register (LLR). Figure 10 shows a pictorial view of the same.

Two scenarios are depicted here. In the first scenario an MCD pings another MCD. A high-level step by step description of the events is shown below:

Step 1: The IP address of the remote MCD gets translated to its Mobile communication device address (this type of translation is usually done by ARP on a network like Ethernet; for simplicity this is assumed to be a static table in our scenario).

Step 2: The IP/ICMP packet gets encapsulated in a packet (with the mobile communication device address of the destination MCD) and sent to the affiliated MBS (using cell multiplexing and SRMA)

Step 3: The source MBS identifies the MBS connected to the destination address using the mobile communication protocol (from the GLR).

Step 4: MBS translates the destination MBS address to an MCA address (a static mapping) and passes the message to its MCA node.

Step 5: MCA node sends the message to destination MCA node (mobility is hidden from MCA users, in this case, Mobile communication signalling).

Step 6: The destination MCA node passes the message to the destination MBS.

Step 7: Destination MBS sends the message to the destination MCD using cell multiplexing.

Step 8: The Mobile communication layer on the destination MCD passes the message up to the IP layer.

Step 9: The same procedure in reverse gets the message back to the source.

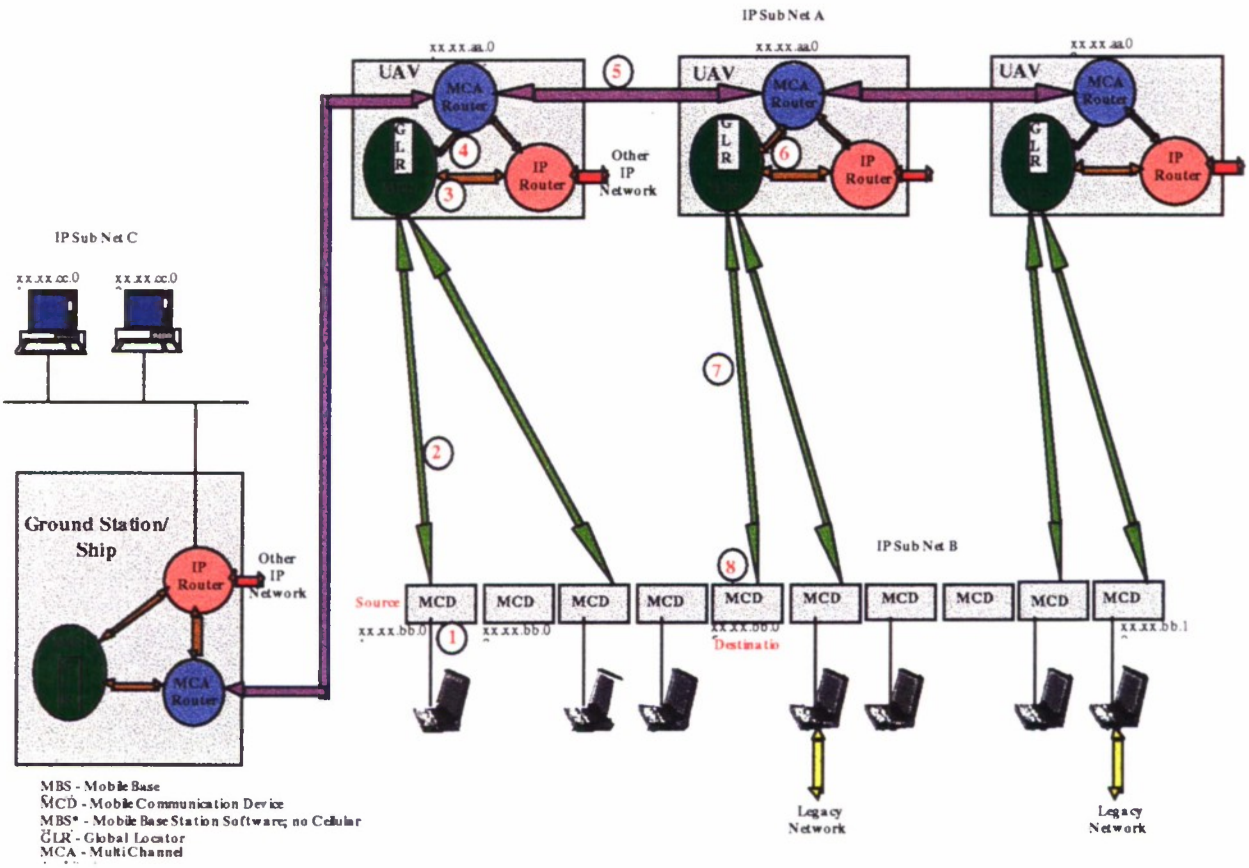


Figure 10. One MCD pinging another MCD.

In the second scenario one MCD pings a node connected to an external IP subnet. The destination selected is a node connected to the IP router on the ground station/ship. A step by step description of events and pictorial view (Figure 11) is shown below. Step 1 begins with the identification of a gateway router. This identification can be managed by dynamically updating the MCD routing table to always use the IP address of the router's mobile communication interface (i.e., the one that is attached to the MBS with which the MS is currently affiliated) as the default gateway. This dynamic updating is not required, but it does make the WI operate more efficiently. Moreover, it is easy to implement since all the MBS/router interconnections are static and, therefore, easily known and the MS always knows which MBS it is currently affiliated with.

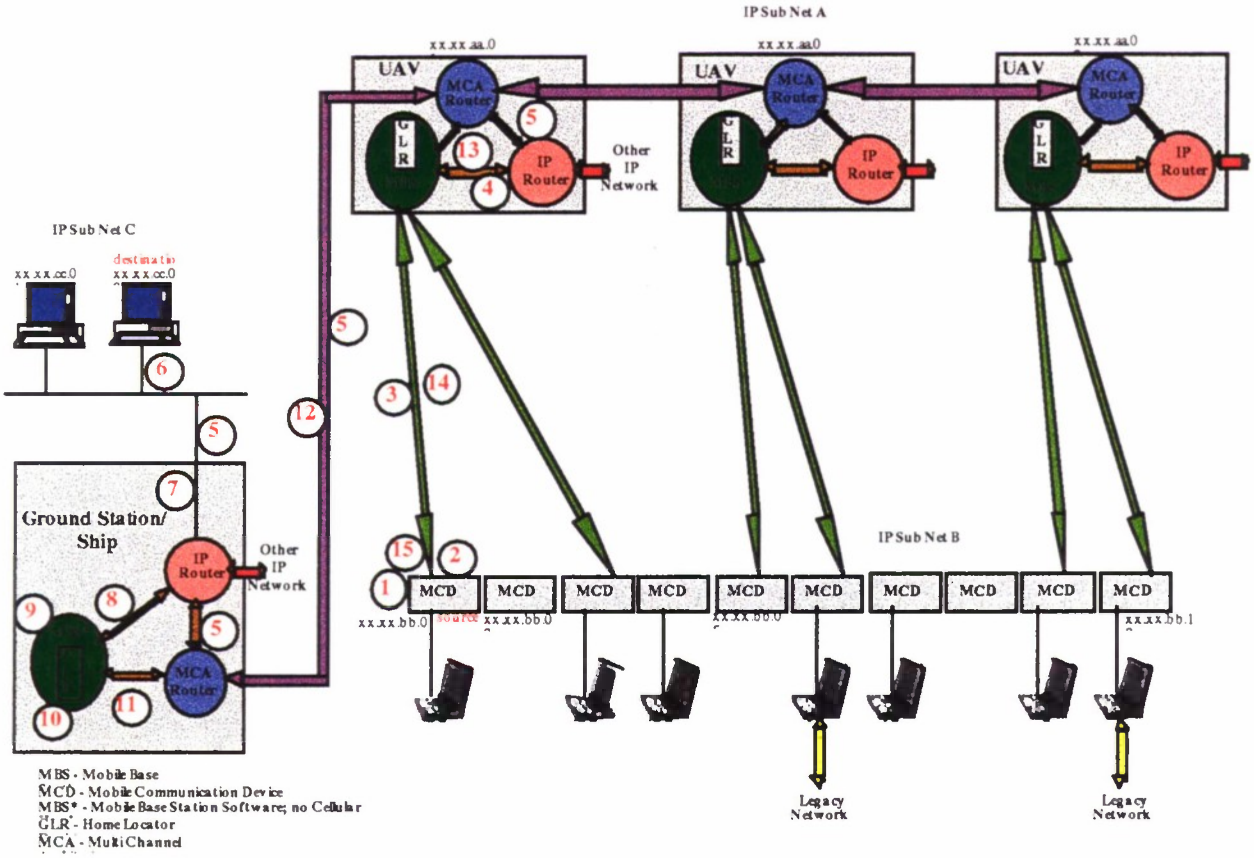


Figure 11. An MCD pinging an external IP node.

Step 1: MCD identifies the gateway router from the mobile communication network (it is the IP router on the airborne platform connected to this particular MCD).

Step 2: The IP address of the gateway router's mobile communication interface is translated to its Mobile communication device address (initial assumption: using a static table).

Step 3: The IP/ICMP packet gets encapsulated in a packet (with the mobile communication device address of the gateway router) and sent to its MBS at the airborne platform (using cell multiplexing and SRMA).

Step 4: The source MBS passes the mobile communication frame to the mobile communication interface of the IP router.

Step 5: The IP router routes the IP packet (from the mobile communication packet) to its final destination using MCA.

Step 6: In reverse, the destination IP node identifies its gateway router to the mobile communication network using the ARP protocol.

Step 7: The IP/ICMP packet is sent to the gateway IP router.

Step 8: The gateway IP router translates the source (of the ping) IP address to its mobile communication device address (initial assumption: using a static table) and transfers the IP packet and destination mobile communication device address to MBS*.

Step 9: MBS* identifies the mobile communication device address of the MBS connected to the MCD (the source of the ping) by consulting with the GLR.

Step 10: MBS* then identifies the MCA router (mobile communication device—MCA address translation; initial assumption: static) connected to the MBS, which in turn is connected to the source of the ping (MCD).

Step 11: MBS* then encapsulates the IP packet in a mobile communication device frame and passes it along with the destination MCA address to its corresponding MCA router.

Step 12: The mobile communication device frame is sent to the respective MCA node using MCA protocol.

Step 13: The MCA node passes the mobile communication device frame to the corresponding MBS.

Step 14: MBS sends the message to the source MCD using cell multiplexing.

Step 15: The mobile communication layer at the source MCD passes the message up to the IP layer.

References

1. Dennis J. Baker, James P. Hauser, Dennis N. McGregor, and James T. Ramsey, "The UNT/NRL HF Inratask Force Communication Network Experiment," NRL/MR/4440--92-6965, June 4, 1992.

2. James P. Hauser, "Voice Management and Multiplexing Protocols Developed for the Data and Voice Integration Advanced Technology Demonstration," NRL/MR/5521--95-7792, November 13, 1995.

3. William A. Thoet, Dennis J. Baker, and Dennis N. McGregor, "A Multichannel Architecture for Naval Task Force Communication," NRL/FR/5520--94-9703, January 30, 1994.

APPENDIX C
SHARED RESOURCES MULTIPLE ACCESS (SRMA)

Shared Resources Multiple Access (SRMA) is a methodology for dynamically allocating uplink RF receiver media resource access among a large number of multiple users based on actual traffic loading. The ability to provide integrated services with QoS commitments to WI mobile subscribers is supported on both the uplinks and downlinks by the Integrated Services Cell Multiplexing (ISCM) protocol. The ISCM is responsible for offered traffic loading information, Quality of Service (QoS) commitments, integration of voice and data into a single stream of cells (data stream), reservation of network capacities to support Virtual Circuits (VC) via RSVP and the delivery of datagrams. However, since many mobile subscribers must access a common set of uplink RF resources, the WI needs an efficient Medium Access Control (MAC) Layer protocol that also supports QoS commitments for managing the uplinks. Only those functions of the ISCM directly related to it will be included here in order to present an overall understanding of SRMA by presenting an actual implementation.

SRMA provides the mechanisms to fully maximize access to the uplink RF receiver resources available at a base station using in-band signalling. A Master Transmission Scheduler (MTS) located within the Base Station Controller (BSC) uses a host of independent decision making modules to ensure that maximum utilization of uplink RF resources is realized while at the same time minimizing the overhead imposed on the downlink data stream. This is accomplished by two-dimensional scheduling of uplink access based on both time and code/frequency. Some of the modules required by the MTS are data source specific and the model presented here is for an ISCM implementation only.

1. SRMA Features List

- Compatible with proven cell multiplexing scheme for managing integrated services.
- Ability to schedule on a two-dimensional basis (time/code).
- Dynamic frame-by-frame uplink scheduling allowing quick and efficient management of changing traffic loading and user response times.
- Provisions for multiple transmissions per frame.
- Flexible bid request scheduling with provisions for variable position in frame, variable frequency/code used and multiple opportunities per frame.
- Contention for uplink access occurs only during announced bid opportunities minimizing collisions.
- Standard COTS bid contention protocols supported.
- Scheduling performed in centralized airborne nodes providing single point for upgrades.
- Scheduler designed to be modular allowing additional capabilities to be added gradually.
- Inherent automatic ranging control.
- Low scheduling requirements mean low probability of scheduling errors.
- Effective error handling.
- In-band signalling reduces hardware assets and power requirements.

- Designed for maximum possible efficiency allowing more capable hardware and software to be added as they become available without need to modify protocol or existing software.
- Protocol supports built-in paging function.

2. Description of SRMA

The downlink traffic broadcast by the BSC is broken into variable length frames, each consisting of a unique Beginning of Frame (BOF) identifier that contains a unique BSC identifier, a list of Mobile Communications Device (MCD) uplink schedules, a unique Beginning of Data (BOD) identifier, and broadcast data. The MCD is the hand-held communication terminal that hosts an MS node. BSC uplink RF receiver resource access is granted by the BSC on a frame-by-frame basis by means of uplink transmission scheduling. The results of the MTS scheduling algorithms are broadcast down to all MCDs and used by their Slave Transmission Controllers (STC) to regulate the time intervals within the uplink frames that their transmissions are allowed to occur. Included in the list of schedules are Bid Solicitation Messages (BSM) that provide an uplink transmission opportunity to all inactive MCDs to announce their request for inclusion in the access scheduling by sending their first Traffic Transmission Requirement (TTR). The actual scheduling of BSM may occur anywhere in the current frame and may occur more than once per frame. This provides flexibility for both security reasons and improved response times. Bids that have been accepted from the previous frame are announced via Bid Acknowledgment Messages (BAM) and are followed by a first transmission schedule for each newly active MCD. The SRMA protocol provides an addressing mode for all commands to the MCD. The addressing modes supported are individual MCD (Unicast), group of MCDs (multicast) and all MCDs (broadcast). The BSC identifier within the BOF is used by the MCD to determine with which BSC it is currently in communication. When it is in an active state, the MCD can thus detect when a change occurs in BSC coverage and initiate a handover to a new BSC.

The uplink bandwidth is divided into variable length frames by the MTS. MCDs, when active, are allocated slots of time within each frame to transmit based on their TTR from the previous frame or initial bid requests. At each subsequent allotted time an MCD will transmit its then current TTR along with the data that was scheduled during the preceding frame. If no data was scheduled for the current frame, then only the TTR will be sent. A Repeat Schedule Mode (RSM) option allows the MCD to continue transmitting on the same schedule until a new schedule is received. This feature is used to minimize downlink overhead when communications quality is high and conditions are relatively static. When no data is available to be sent in the next frame, the MTS has the option to reallocate that time slot to another MCD that does have data available. This allows reserved but unused capacity to be distributed to other active MCDs on a frame-by-frame basis, thereby increasing overall uplink throughput while still maintaining individual QoS.

Bid Controller (BC) modules in both the BSC and MCD control all access to system resources. When an inactive MCD detects that new user input data is available, it will monitor the BSC downlink frames for a BSM. At the next uplink bid opportunity, a Bid Request Message (BRM) will be sent to the BSC. The BSC BC will perform all necessary functions required to ensure that the MCD is authorized to access the system. In addition to simple unit authorization, availability of resources requested will be verified. If all conditions are met, the MCD is assigned a temporary local user identification number for SRMA control. Registrations with the network Domain Name Server (DNS) and network Foreign Agent (FA), etc., are accomplished as higher level functions. When a change of BSC (or a change of beams within the same BSC) is detected while an MCD is active, a Service Transfer Message (STM) will be transmitted to the new BSC at the time BSM are authorized. The STM will instruct the current BSC BC to effect a transfer of all services from the previous BSC association to the new BSC association. When all users have signed off or there has been no data available from an active user for a period of time (specified in the Complan), the MCD will terminate its current session, becoming inactive once again.

It should be noted that SRMA can function equally effectively with integrated services protocols other than ISCM, such as ATM, provided that those protocols supply the traffic loading information (either on a known "traffic in queue" or predictive basis), the required MTS scheduling modules, and Bid Controller interfacing. Additionally, a time slice of the uplink frames could be reserved by SRMA for use by other protocols, such as ALOHA, should such a need ever arise. Additional functionality would have to be added to the BSC and MCD to support the additional protocol. However, it should be noted that using this approach may place restrictions on SRMA that might impede its ability to schedule most effectively.

3. Components

3.1 Base Station Controller (BSC)

The interface between the BSC and the Mobile Router (MR) is described in Appendix B. Functionally, the BSC contains the modules described below and resides in either a WI airborne or terrestrial Mobile Base Station. Its responsibilities include the formulation of uplink transmission schedules and the multiplexing of cells queued in the Service Ports onto the BSC downlink (ISCM) (see Figure 1).

3.2 Multiplexer

The ISCM Multiplexer receives data from the Mobile Router (MR). The data flow contains Cellular/PCS destination address information as well as service identification information that permits the data to be handed to the proper ISCM Service Port.

3.3 De-multiplexer

The ISCM De-multiplexer receives multiple data streams from the SRMA decoder and sends them to the MR. The datagram and VC Signalling virtual channels are unpacked; VC cells are passed through without modification because they are transported end-to-end.

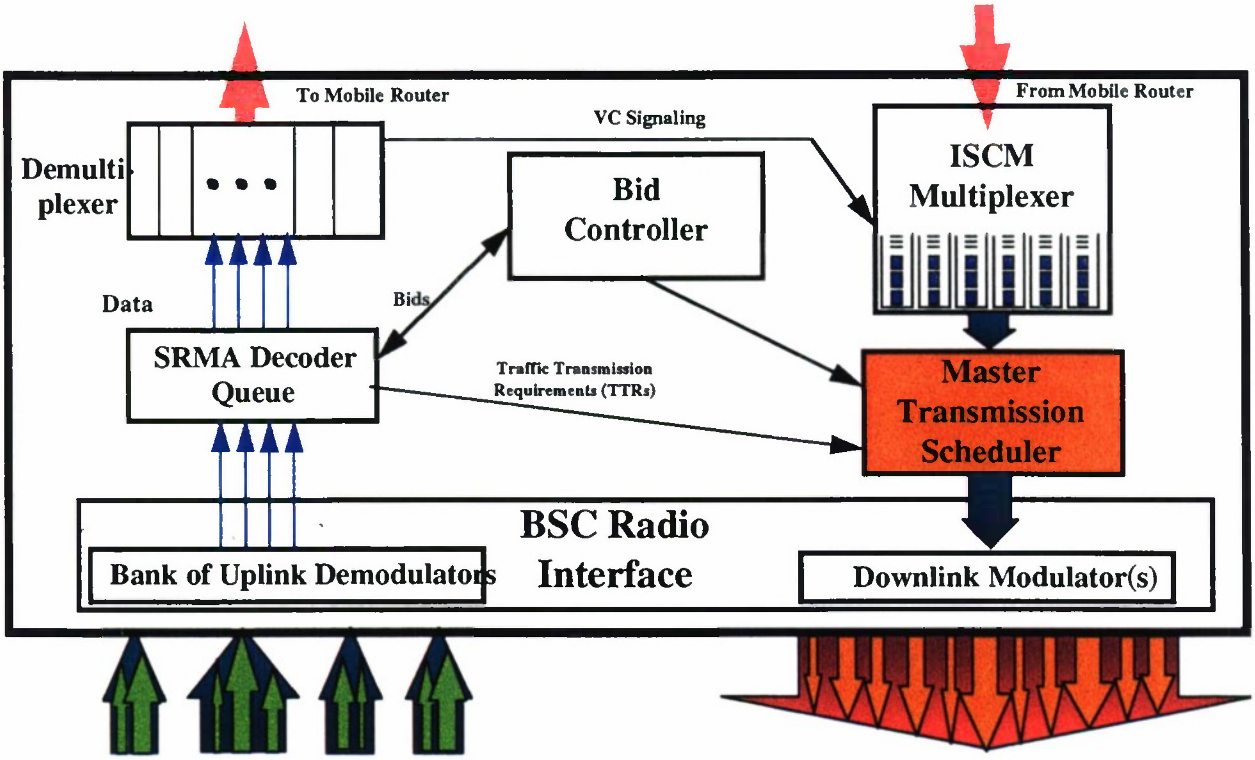


Figure 1. Base Station Controller.

3.4 SRMA Decoder

The BSC SRMA Decoder is responsible for monitoring the multiple RF uplink channels and decoding the SRMA protocol received from the various active MCD units. Its main function is to separate and route imbedded commands and data. Commands are separated into bid and TTR components and distributed to the bid controller and MTS, respectively. Data is passed to the de-multiplexer for processing.

3.5 Bid Controller

The BSC Bid Controller (BSCBC) is responsible for registering an MCD for access to the system uplink RF resources. When a bid request is received from an inactive MCD, its hardware identifier (or other unique identifier) is passed on to an authentication center to ensure that it is eligible to participate in the Cellular/PCS Subnet. Once that has been successfully accomplished, the MCD is attached to the subnet by performing whatever registration, etc., is required to properly log the device on as an active participant. This includes assigning a local user identification number for SRMA control and possibly running ARP (Address Resolution Protocol) to disseminate the mapping between the Cellular/PCS Subnet address and the IP address. Finally, a bid acknowledgment message is sent back to the MCD as part of the next SRMA frame. This informs the MCD that its bid has been accepted and assigns the MCD a temporary session user ID number. The MCD is then put into an active state and begins monitoring the downlink for its individual uplink schedule. The BSC BC is also responsible for effecting the handover of an active MCD from a previous BSC upon receipt of an STM.

3.6 Master Transmission Scheduler (MTS)

The Master Transmission Scheduler (MTS) has the overall responsibility for assigning the available RF receiver uplink resources to multiple subscribers in the most efficient manner possible (see Figure 2). It uses a host of independent decision making modules to ensure that maximum utilization of these resources is realized while at the same time minimizing the overhead imposed on the downlink data stream. By carefully balancing these two main requirements, the MTS supports as many subscribers at possible at any given moment in time. This is accomplished by the frame-by-frame scheduling of media resources based on the current ISCM (or other multiplexing layer) traffic requirements. Each active MCD will send a TTR message during each frame accompanied by data scheduled from the previous frame. When an active MCD indicates that it does not intend to transmit during the next frame time, its normally allocated time may be reassigned to another MCD that does have traffic to send. The decision support modules include Complan, Ranging Control, System Loading, Resource Allocations and Traffic Analysis, which are common to all types of data streams. Data stream (i.e., multiplexing layer protocol) specific modules include TTR Control, Bid Control and Priority/precedence. The MTS is designed so that the modules are independent of the MTS and each other and communicate via a standard Application Program Interface. Modules may be added in any order and updated as deemed necessary to increase functionality without requiring changes to other BSC functions, the MCD, or the SRMA protocol This provides a highly effective methodology for systematic upgrading of the MTS from the most basic of scheduling techniques, such as TDMA, up to very complex algorithms providing the utmost in scheduling efficiencies.

3.7 BSC Radio Interface

The BSC Radio Interface (BSCRI) provides the interface between the SRMA software modules and the RF uplink and downlink hardware. The BSCRI hardware may control either a single beam or a multiple beam array.

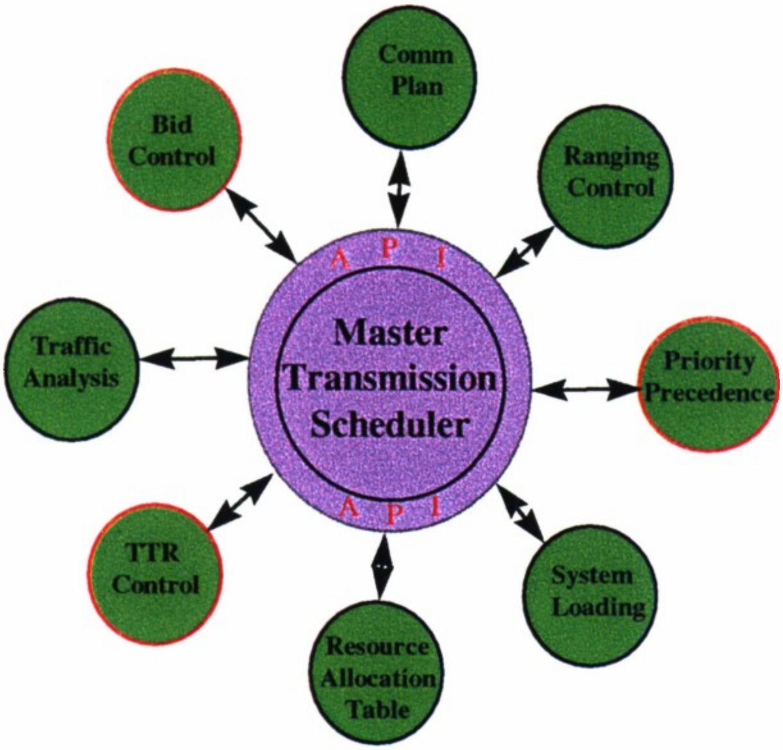


Figure 2. SRMA Master Transmission Scheduler.

4. Mobile Communications Device (MCD)

The MCD is a unique hand held RF radio device that contains special hardware and software for interaction with a WI BSC (see Figure 3). The MCD also provides the software interfaces between the user applications and the SRMA protocol. Functionally it contains the modules described below and may be optionally connected to a Note Book Computer (NBC) which may in turn be connected to a wired or wireless LAN. User Applications may be located within the MCD itself or externally via an NBC RS232C interface. Its responsibilities include integration of voice and data into a single stream of cells (ISCM) transmitted to the BSC under control of the STC and obtaining access to BSC resources via the SRMA bidding process.

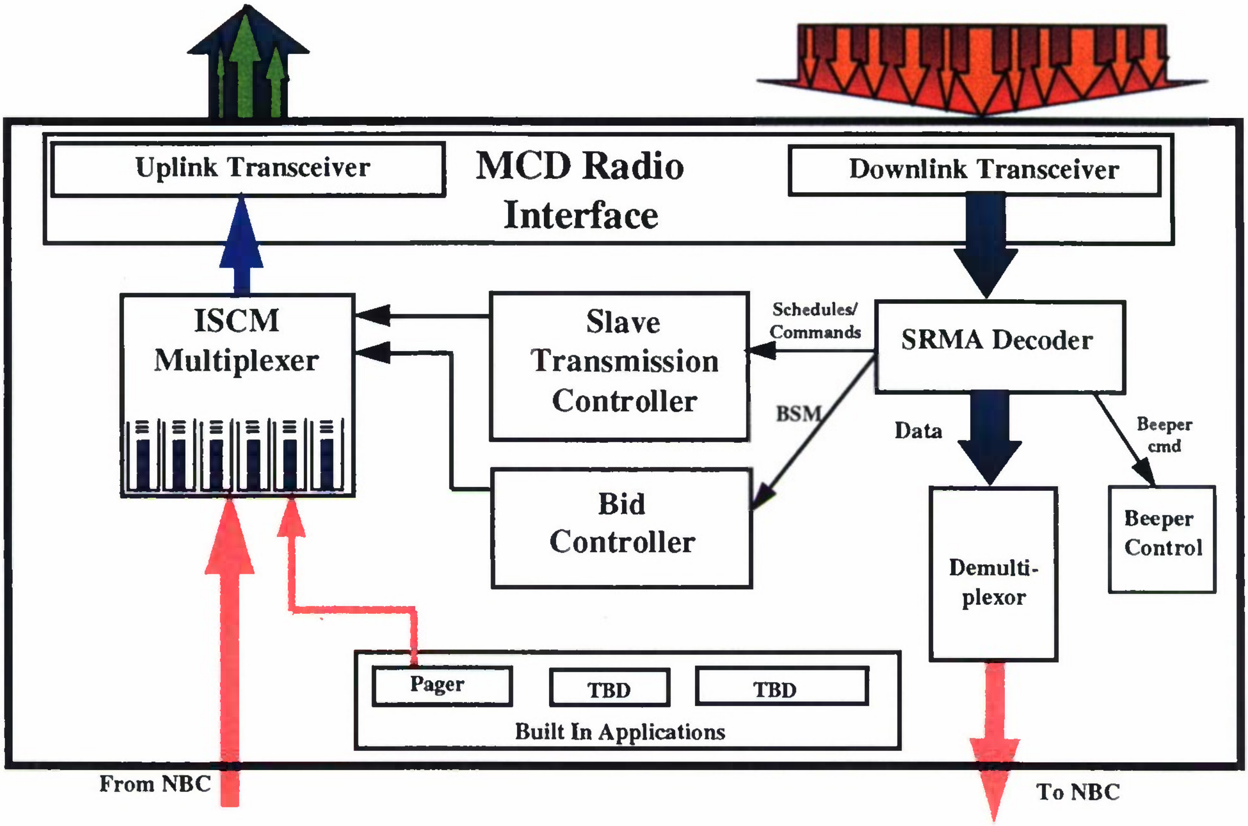


Figure 3. SRMA Mobile Communications Device (MCD).

4.1 Integrated Services Cell Multiplexer (ISCM)

The ISCM Multiplexer determines the order in which cells are transmitted on the uplink to the BSC (see Section 4 for details). IP traffic from local users arrives at the MCD SNPI. By the time an IP flow arrives at the SNPI, it has already been differentiated by the RSVP Classifier function and, therefore, associated with a particular service and ISCM Service Port.

4.2 De-multiplexer

The ISCM De-multiplexer receives the broadcast traffic stream from the SRMA decoder and distributes it to the addressed users. The de-multiplexer unpacks both Cell Layer and Adaptation Layer payloads in order to recover the IP packets contained in those payloads.

4.3 SRMA Decoder

The MCD SRMA decoder is responsible for monitoring the downlink channel and decoding the SRMA protocol to separate and route commands and data. Commands are separated into bid and uplink transmission schedule components and distributed to the BC and Slave Transmission Controller (STC), respectively. Data is passed to the de-multiplexer for processing.

4.4 Bid Controller

The MCD Bid Controller (MCD BC) is responsible for registering an MCD for access to the system RF resources. This is accomplished by monitoring the RF downlink for BSM commands. These messages announce that the BSC will accept new bid requests at a predetermined time slot within the next frame. If new user data becomes available while an MCD is inactive, then a bid request will be sent at the next announced time. The bid request will consist of the unit's unique ID and its first TTR. The MCD will continue to respond to all BSM until a BAM is received assigning a temporary session user device ID. At this point the MCD is placed in the active state and begins monitoring the downlink for its uplink transmission schedule. If active communications with a BSC (or beam within the BSC) are lost or conditions warrant a switch to a higher quality link, the MCDBC will transmit an STM to the newly desired BSC in response to its BSM. The MCD will continue to transmit STM in response to all BSM until a BAM is received assigning a new temporary session user device ID. At this point the MCD is returned to the active state and begins monitoring the downlink of the new BSC for its new uplink transmission schedule.

4.5 Slave Transmission Controller (STC)

The Slave Transmission Controller (STC) has the overall responsibility for controlling the time at which uplink transmissions occur. It is driven directly by the BSC MTS and does no scheduling of its own. Provision is made so that each active MCD will send a TTR during each assigned frame's transmission window accompanied by data scheduled from the previous frame.

4.6 MCD Radio Interface

The MCD Radio Interface (MCDRI) provides the interface between the SRMA software modules and the RF uplink hardware. The MCDRI provides a single RF path for multiplexed traffic to be transmitted to a BSC. A built-in scanner is used to scan all WI BSC and determines which are suitable for use. This information is used by the MCD BC to select a BSC to register with and to transfer to a different BSC when conditions so warrant.

5. Protocol Formats

5.1 Downlink

The downlink traffic broadcast by the BSC is broken into variable length frames each consisting of a unique Beginning of Frame (BOF) identifier, a list of MCD uplink schedules, a unique Beginning of Data (BOD) identifier and ISCM cells (or other data if a different multiplexing layer protocol is used) (see Figures 4 and 5).

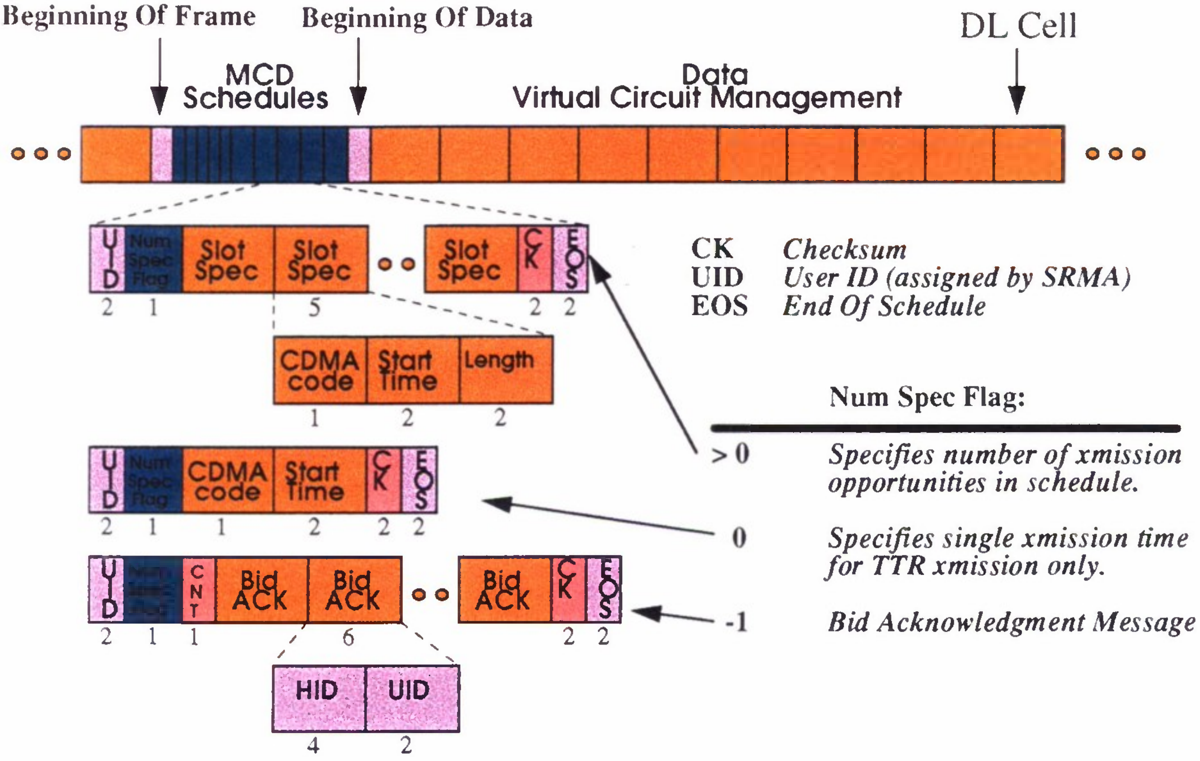


Figure 4. SRMA downlink frame format and MCD scheduling message formats.

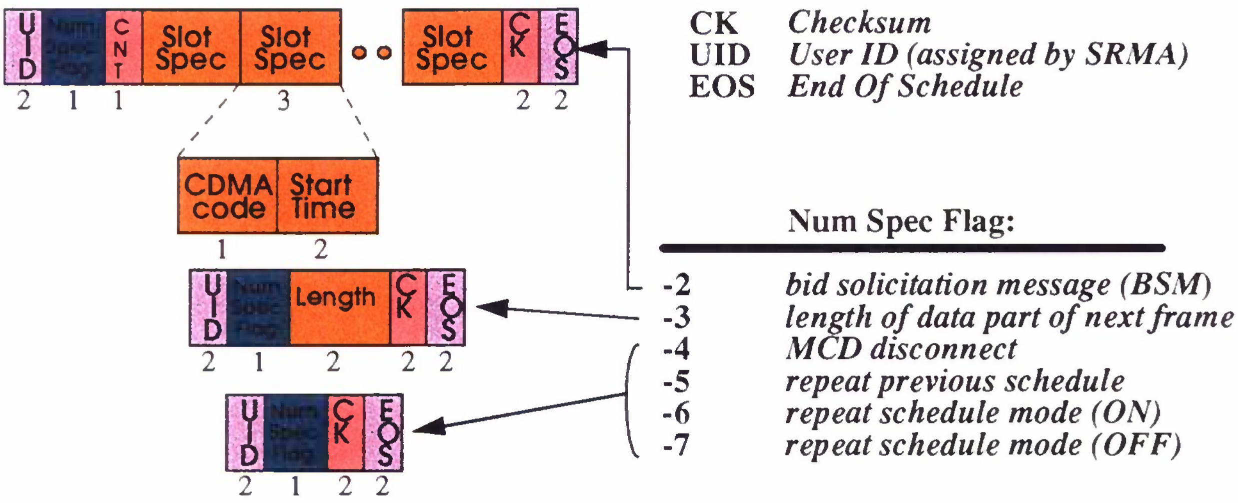


Figure 5. Additional control message formats sent in the MCD schedules part of a downlink frame.

5.2 Uplink

The uplink bandwidth is divided into variable length frames by the MTS (see Figure 6). Subscribers, when active, are allocated a slot of time within each frame to transmit based on their TTR from the previous frame or initial bid.

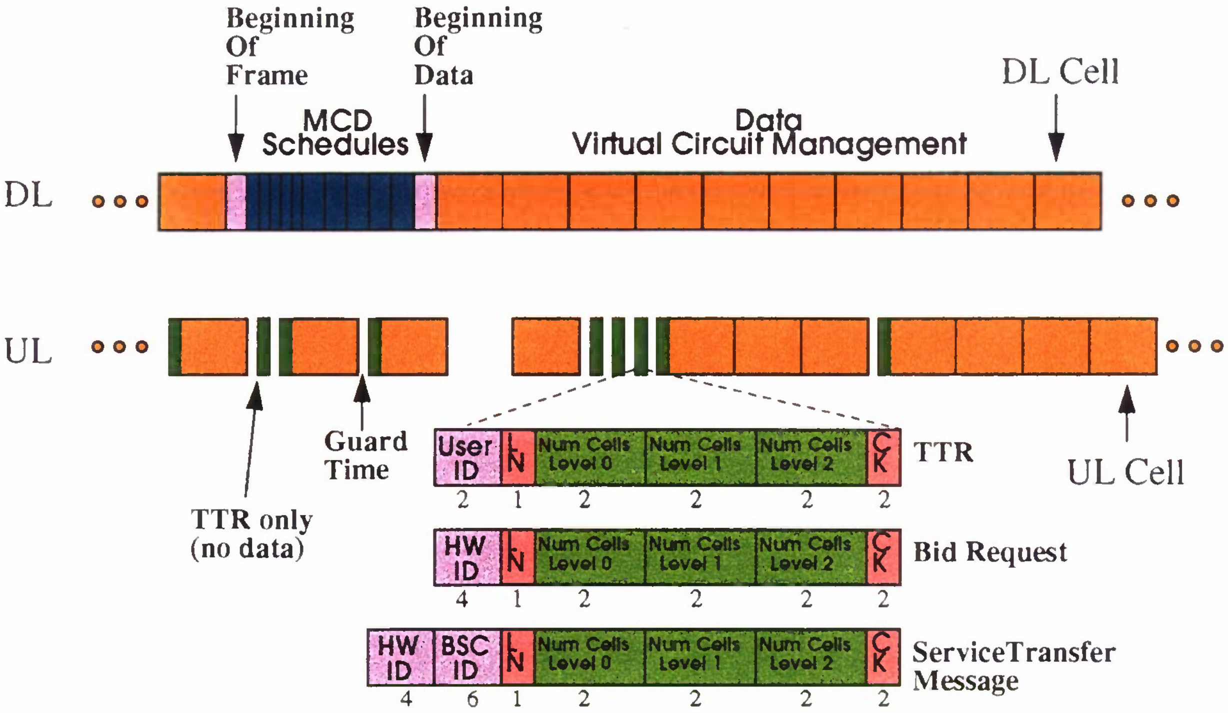


Figure 6. SRMA layer protocol uplink frame.

6. Error Handling

The simplicity of the SRMA environment provides for robust recovery from error conditions usually within one frame time. The use of the repeat schedule mode further reduces the impact of errors on the downlink broadcast.

What happens when the transmission schedule is hit on the downlink?

- Repeat Schedule Mode:
  - Loss of schedule change creates potential for collision.
  - Collision detected by MTS and affected MCD rescheduled.
  - Recovery within as little as one frame time.
- Non-Repeat Schedule Mode:
  - MCD doesn't transmit current frame.
  - No potential for collisions.
  - MTS detects loss of data/TTR; can compensate in next frame.

What happens when the TTR is hit on uplink?

- Repeat Schedule Mode:
  - Minimum effect; bandwidth change delayed one frame.
- Non-repeat Schedule Mode:
  - MTS has option to repeat schedule or schedule TTR only.

7. Glossary

A

ARP	Address Resolution Protocol

B

BAM	Bid Acknowledgment Message
BC	Bid Controller
BLOS	Beyond Line of Sight
BOD	Beginning of Data
BOF	Beginning of Frame
BRM	Bid Request Message
BSC	Base Station Controller
BSCBC	BSC Bid Controller
BSCRI	BSC Radio Interface
BSM	Bid Solicitation Message
BSS	Base Station System
BTS	Base Station Transceiver System

C

CDPD	Cellular Digital Packet Data

D

DG	Datagram

F

FIFO	First In First Out

G

GLR	Global Location Register
GPRS	General Packet Radio Service
GSM	Group Special Mobile

I

IP	Internet Protocol
IS	Integrated Services
ISCM	Integrated Services Cell Multiplexing

L

LLR	Local Location Register

M

MBS	Mobile Base Station
MCA	Multi Channel Architecture
MCD	Mobile Communications Device
MCDBC	MCD Bid Controller
MCDRI	MCD Radio Interface
MR	Mobile Router
MRS	Mobile Routing System
MS	Mobile Subscriber
MTS	Master Transmission Scheduler

N

NBC	Note Book Computer

Q

QoS	Quality of Service

R

RARP	Reverse ARP
RSM	Repeat Schedule Mode
RSVP	Resource Reservation Protocol

S

SNPI	Subnet Provider Interface
SRMA	Shared Resources Multiple Access
STC	Slave Transmission Controller
STM	Service Transfer Message

T

TTR	Traffic Transmission Requirement

V

VC	Virtual Circuit

APPENDIX D
ATM INTEGRATION FOR ARCHITECTURE 2

There are two types of ATM network traffic that may need to traverse the WI network. In the first case, the ATM traffic is destined for an ATM node outside the WI network but needs to use WI connectivity to get to the final destination. In this scenario ATM traffic will be tunneled through the WI network to the exit point. ATM cells will get a small Subnet Layer wrapper for traversing the WI network (see Figure 1). The wrapper is attached at the entry interface box and then is pealed off at the exit interface box. The interface box may be similar in concept to the Quality Controlled Bi-Level Multicast Router (QCBMR) box used to tunnel IP traffic through an ATM network. The second type of ATM traffic is the IP over ATM kind destined for a WI end user. In this scenario the IP packets are extracted from the ATM cells at the WI network entry point interface box and then traverse the WI network to the WI end node.

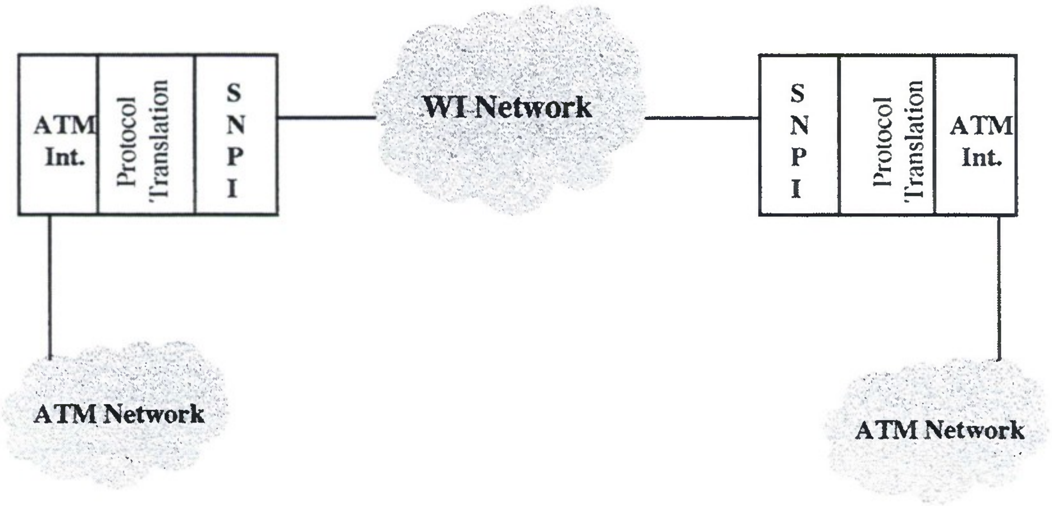


Figure 1. WI Architectural Option # 2 ATM integration.

APPENDIX E
WI ARCHITECTURE DESCRIPTION (OPTION 1)

1. Overview

The Option #1 Architecture is designed with the understanding that data is going to be the dominant form of information. This is not to dismiss voice use but there are many known ways to handle voice reasonably efficiently and the recommendations on handling voice are straightforward. Our principal design goal is to better match the characteristics of data communications in the aggregate to available communications bandwidth. This can improve the efficiency of data channel utilization by orders of magnitude. Under this assumption, Architecture #1 closely follows the expected evolution of the global Internet and consequently the proposed network will be based on the evolving IPv6.

If voice were to be the major information type, it is clear that the standard circuit-oriented cellular-based systems would be more appropriate, and these could form the basic architectural structure on the basis of what is commonly provided today. Alternatively, it may be feasible to exploit the newer fixed packet size systems (like ATM) which still provides the efficiency of statistical multiplexing but will accommodate both isochronous (such as voice) and non-isochronous data traffic with different grades of service. While ATM is clearly appropriate for high bandwidth, low loss fiber channels, it remains to be shown that it has advantages operating over point-to-point dynamic data links, on multidrop broadcast radio links, and on multiple access radio links.

However, it should be noted that for all current data applications, there is an underlying assumption of TCP/IP support. IP will thus be common for all architectures so there is a basic IP internetworking commonality. As a result of the expected uplink data rates and the use of low rate (2400 bps) voice codecs, there are a number of similarities between Architectures #1 and #2 and even some commonality with some of the anticipated ATM wireless technologies. However, there are still considerable differences in implementation. Since there are yet no specific wireless ATM proposals, any comparisons have to be limited to differences between #1 and #2. There are significant differences in the subscriber subnet portion with respect to multiple access techniques. In both cases within the subscriber subnet there is a segmentation in fixed size transmission units to accommodate voice needs. Another difference in architectures is that in Architecture #1, the backbone network is assumed to be pure IP with variable size packets and using IP technology that accommodates isochronous traffic. The goal was to incorporate commercial routers with enhanced software accounting for the mobility of the backbone.

In contrast, Architecture #2 proposes a backbone switching structure that still transports fixed size transmission units. This requires a unique backbone switch development. Architecture #2 will evaluate the merits of the fixed packet (cell) systems for the entire mobile WI. This architecture will, in fact, address many of the same issues that wireless ATM would have to face in the future. In the Architecture #1 internetworking (also shortened to interworking) considerations, there is recognition that ATM subnets will be part of the larger network of networks. The data interworking is still IP-based with the assumption that the different subnets **will be** a mixture of ATM and non-ATM networks. However, the WI subnet of Architecture #1 is not ATM cell-based!

There needs to be a clear warning concerning the implementation suggestions offered in this appendix. We have risked making many design implementation assumptions so as to highlight many of the next level down system level considerations that would not have been raised if we had remained at a very general but high level. The tradeoff space is so multidimensional in a complex unprecedented system like the WI that one is forced to offer a strawman solution to a moderate level of detail based not on detailed analyses but rather on a synthesis based on somewhat related systems that have been built, plus a set of heuristics based on experience. Once the fundamental framework has been synthesized, then one can perform detailed analyses on various aspects of the system. One can produce point optimizations that maximize some performance parameter but each of these optimizations have to be measured against complexity increases in some other dimension.

As one example of the trade space complexity, we have proposed 32 uplink demodulators serving each beam. Users certainly may independently generate requests for communications with a distribution that could approximate a Poisson process or could be clustered in a time sense depending on the tactical situation. One has to provide some scheme where the user is offered uplink capacity rights on one or more of the demodulators. In the offered scheme of Architecture option #1, we have proposed that the "system" will assign a user to share a single uplink demodulator for the duration of a particular set of transactions and this assignment is made depending on the type of traffic and the priority of the user. This may not be one's first choice. We recognize that if a group of users is not restricted to a single demodulator but had access to many demodulators, the overall efficiency of utilization could be improved. One knows intuitively that given a completely fair service policy, the optimum process would be to let a particular user's uplink transaction entities be spread among all the demodulators based on the dynamic occupancy of the individual channels.

Even though the ability of a user's uplink session to access multiple channels is potentially more efficient, we believe that many of the attributes of single channel assignment (other than efficiency) may swing the decision in the favor of the latter. When users have different access priorities and messages have different precedence, a user partitioning (enabled by the subscriber channel controller) may be very useful and simpler to implement. In any case, this is the going-in assumption on architectural option #1.

Let us give a more pragmatic example where a better solution in term of communications efficiency may not be implemented for reasons other than purely technical ones. In many defense-oriented organizations, one could use a single 100 Mbps Ethernet service rather than say ten or less individual 10 Mbps Ethernets bridged together. The 100 Mbps would certainly be more efficient in a traffic sense. However, there may be strong security reasons to compartmentalize a number of 10 Mbps Ethernets to service specific classified programs. While the technology may be available to provide sufficient security at the end user, the logistics and operation of the security solution may be deemed inadequate. Similarly, there may be a perception in the graphics design department that they "need" a dedicated 10 Ethernet since they know that they have a high load and don't want to suffer interactive delays due to unpredictable loads from the other parts of the organization. In a complex system design, sometimes both rational and less-than-rational user perceptions have to be accommodated. Returning to the WI uplink situation, we believe that efficiency must be a key criteria but beyond a certain level of improvement other factors deserve equal attention. This claim will be clarified as the Architecture #1 strawman is expanded on in this appendix.

IP has always been a "best effort" service. Packetized voice services have occasionally been demonstrated over the past 20 years, but it has always been acknowledged that a highly loaded network will not provide an acceptable level of voice service because delay and delay variability are not controllable. This will no longer will be true when IPv6 is implemented. Consequently, Architecture #1 will follow the IPv6 direction. Note that voice packetization will still tend to smaller packet sizes consistent with limiting the overall delay for multiple router relays. The lower limit on the packet size is conditioned by the larger percentage of overhead levied by the larger packet headers (compared to cell headers).

In the mobile WI we have two major subsystems (the subscriber subsystem and the backbone subsystem) and an auxiliary mobility support subsystem which is closely affiliated with providing subscriber related services. The focus of this appendix is on the design considerations of both the airborne portion of the subscriber subsystem and the subscriber's terminal (MCD).

In the option #1 architecture, we have postulated two separate routers: a subscriber router function which locally handles some of its end point connection changes in a network-wide transparent manner and the backbone router which is a more traditional IP router augmented to support IPv6. We will try to keep these routers separated as long as possible and only when each is specified sufficiently will we see if it makes sense to combine them into a single physical subsystem. Figure 1 is our strawman architecture for the airborne node which is used to identify the overall relationship between the functional subsystems.

In the section on basic operations, we will examine some of the operational aspects of the airborne system functional block interactions so as to give some motivation for more detailed functional block discussions.

It is useful to point out at this point that a user (after authentication and following a request for a particular class of information transfer) is associated with a specific uplink demodulator for the duration of the session. Many different users can share this same demodulator. We initially will consider placing isochronous data (voice) and data into different demodulators since it may prove easier to provide better grades of services to them in this manner. The users sharing the same demodulator will use the same spread spectrum code. Each demodulator will have its own code and hence users in one demodulator will appear to users in another demodulator essentially as "white" noise.

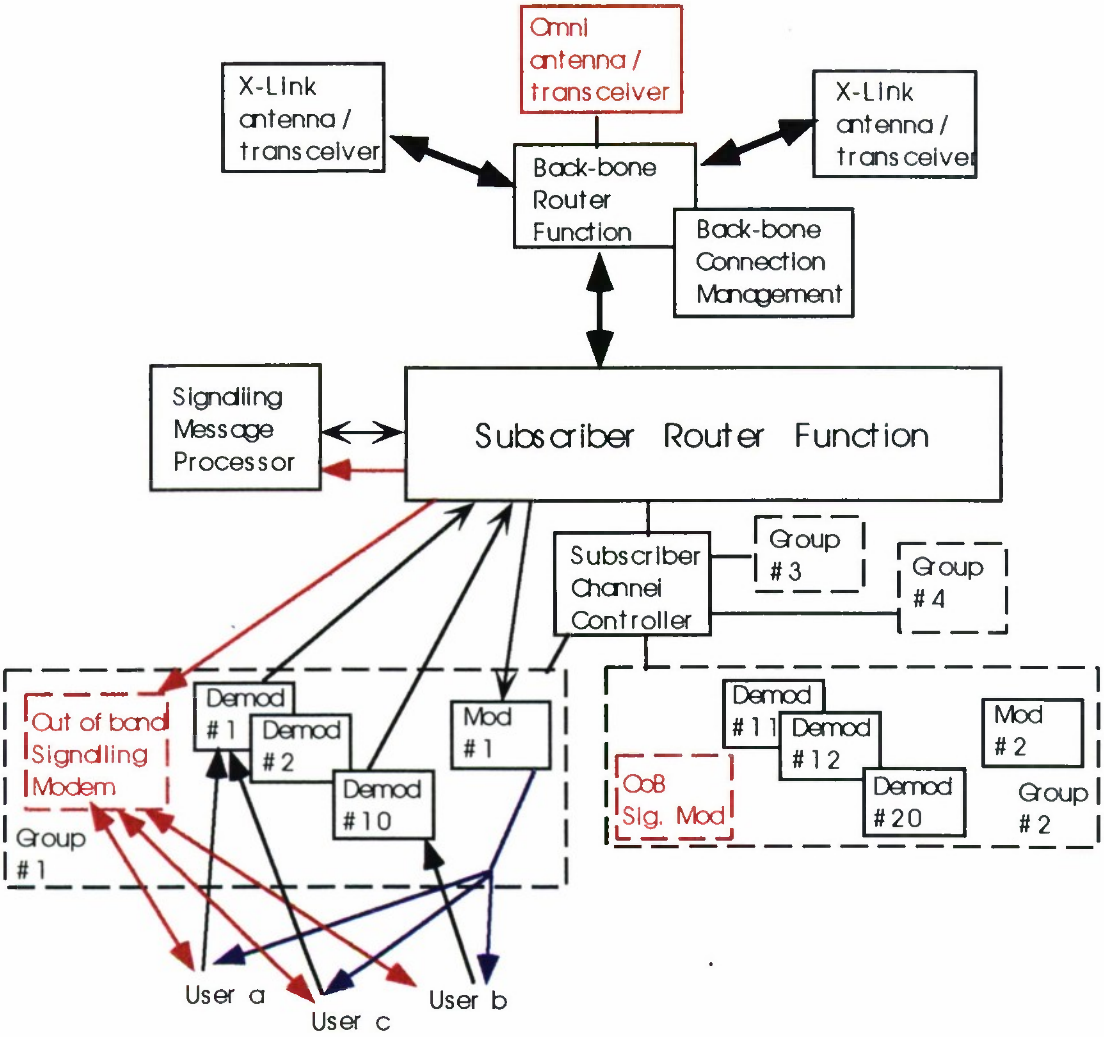

Omni antenna / transceiver
X-Link antenna / transceiver
X-Link antenna / transceiver
Back-bone Router Function
Back-bone Connection Management
Signaling Message Processor
Subscriber Router Function
Subscriber Channel Controller
Group #3
Group #4
Out of band Signaling Modem
Demod #1
Demod #2
Demod #10
Mod #1
Group #1
Demod #11
Demod #12
Demod #20
Mod #2
OoB Sig. Mod
Group #2
User a
User c
User b

Figure 1. Airborne subsystem functional allocation.

It is always difficult to avoid associating a functional allocation with specific implementation choices. However, planning for a framework for implementation design is something that should be considered very early. It is more than a choice of simple hardware platform or even a suite of commercially-available software modules. We anticipate that for a number of subsystems much of the hardware can be "conventional" but much of the software will be either totally unique or require modifications of already developed software. We will propose a specific but flexible development environment so that all interested developers can share the code developed by the group. The intent is to produce all the code in the public domain. Some of the code may be applicable to support standards efforts.

2. Basic Operations

To describe the functionality of each subsystem in the airborne node and its subsystem interactions, it is useful to describe some of the operations needed to initiate communications. We will give an overview of these operations with respect to the participating subsystems, and in Section 3 we will provide more detail on each subsystem. The next section outlines the important process of registration, which is the process that builds a subscriber base for each airborne platform and provides routing information for the system as a whole.

2.1 Registration

At its simplest, registration refers to the association of a mobile user with a specific airborne platform. Since all users may have home networks other than the mobile WI, a key aspect of the registration process is a confirmation that the user is authorized to join the network. Clearly, in a military specific network, it is imperative to restrict network access to those entities that have a certified need to be attached to the network. During a user's stay in the theater of operation, the user will be serviced by a number of airborne platforms, and as the user switches between coverage patterns, another aspect of the registration process must be invoked to update specific databases.

Thus we will use the term registration to include autoconfiguration as well as part of the authentication process. Part of the process is necessarily initiated manually. In the tactical theater, the user is confronted with access to limited communications resources which must be managed on the behalf of all the users. All users are not automatically considered as having equal priorities as they would be on commercial systems. Priorities can be assigned based on a user's relative importance within the tactical organization and/or on a temporary basis contingent on some special critical tactical operation. The key point is that priorities can be associated with the individual on a slowly varying basis. This is almost a quasi-static priority assignment. Some sort of priority system is needed to ration the scarce communication resources.

Along with the assignment of a user to a basic priority level, there may be a secondary level priority depending on the types of traffic requested. The user traffic descriptor needs to be converted into a request for resource requirements. The term "resource allocation" refers to the interaction of a number of databases including the authenticator.

As noted, we want to distinguish between different aspects of registration. We might define an initial registration which implies that a user is arriving from another network and is registering for the first time. The simplest situation would be when deployment to the tactical field is pre-planned and that a planner would have the responsibility for preloading the WI authentication database with the information needed for the users to complete their authentication process on entry into the theater.

If preloading is not the operational mode, then one would need the authentication system on receiving a "registration request" to be able to access a certified remote authentication center which then updates the WI authentication center so that subsequent authentication sessions do not need long distance authentication services. The details of the authentication permits may be considerably different in the local WI since communication constraints need to be more tightly managed in the WI. The local WI authentication record would need to be edited so as to fit the new arrival into an appropriate priority category.

If a user had registered earlier, then incoming pages and calls can be processed. Note pages and incoming voice calls are handled differently. A page does not generally need an acknowledgment and hence there is no need for a returned message. The completion of an incoming call needs the callee to request demodulation assets, and they can only be granted if capacity is available or the callee's priority is high enough to bump a lower priority call.

Airborne platform subscriber beams form movable cell boundaries on the earth's surface. With either multiple airborne platforms or with a single platform with multiple downlink beams, the user will have changing beam coverage depending on the flight history of the platforms. This obviously effects the user's end-link connectivity point and hence the end-to-end routing. The "handover" process is associated with these changes in end-link connectivity. The relationship between handover and re-registration will be covered in a following subsection.

Initial Registration

The normal registration process occurs when a mobile user first logs on and is a prelude to a request for a specific service. The registration process authenticates the user via the authentication register. This sets up the user for the next level of service requests. The following paragraphs go into slightly more detail.

When a subscriber first logs on to a node (airborne or ground), an authentication process is invoked which assigns a priority to a user and allows the user access to specified types of communications. The authentication process is conducted over the uplink's out-of-band signalling channel since it is not desirable at this time to be given privileged uplink communication assets. The out-of-band signalling may be performed on a separate demodulator (and perhaps modulation stream).

> [Important. In cellular-based telephony, the separate concepts of out-of-band and in-band signalling are distinguished. Out-of-band signalling refers to signalling message transfers over channels separate from data channels; in most cases the user channels are not yet allocated. In-band signalling refers to signalling messages associated with an on-going user session and these may be carried on the same channel as user traffic. This separation of signalling types is retained in this appendix for the following reason.
>
> It is highly desirable to limit the damage that an unauthorized user can inflict onto the mobile WI. The first task in getting access to the WI will be via exchange signalling messages with some "well-known" shared asset. "Well-known" means that there is fixed and globally shared spread spectrum code that is "part" of every MCD (perhaps distributed on a "smart card." This shared code is potentially the weakest part of the security architecture. One can limit the damage by allowing this code to only access the out-of-band facility and thus limit the threat

mainly to a denial of service attack. The out-of-band facility is our first "firewall." Only after a subsequent successful authentication process will a user be granted a data channel and, implicitly, an in-band signalling channel both which are better protected with different secure TRANSEC codes. This in-band signalling channel should be free from denial of service attacks.

Thus, the out-of-band channel is treated as if it were physically a different system when it is meant as a different logical system that may share the same physical assets. The exact physical sharing method is yet to be determined.]

In the following few paragraphs we will assume that the user registered, was authenticated, but had not yet requested an active information exchange. This is a period in which this user's network accessibility to others needs to be distributed. The process to be described is only one possible choice out of many as it is meant to flesh out the various interactions that need to be accomplished between subsystems.

Once a user is certified as an authenticated user, the next phase of the registration process is handled by multicast messaging between the authenticator and all the airborne subscriber subsystems that are active. The authentication server will deliver to the subscriber channel controller a list of the communications permissions granted to the authenticated user. The message contains the user's IP number, the communications permissions, temporary priority level, etc. These are necessary for the subscriber channel controller when the user ultimately requests a "data" channel; its requirements need to be balanced versus the already assigned uplink assets. If necessary, the subscriber may be forced to terminate another user's active session if the latter's priority is not high enough.

Once an uplink channel assignment is made, the subscriber channel controller will generate a broadcast message 1) to the local subscriber router, 2) to the "foreign agent" router at the mobile entry node, and 3) to the subscriber routers on the other platforms. This message binds the user's home IP address and its current airborne SRF IP address. The end result of this process will be two databases associated with the subscriber router. The first database is a listing of all the active mobile users attached to this particular platform (local database). The second database is a listing of active users connected anywhere on the mobile WI (a home IP/subscriber router IP tuple global database). The router at the mobile entry node builds the WI specific global database. This router at the mobile entry node which is configured as the "foreign agent" will send a message to the subscriber's home agent. (This is done only if this registration identifies the member as a new entry to the network. The authentication process has to be sufficiently distributed so as to allow a completely new member to the WI not to be rejected. This is a manual planning process completed before deployment.)

The registration process, through its broadcast process, has allowed every node on the backbone to know how to reach the registered member via its local subscriber router. This is important since it is useful to know how to reach a user for incoming calls or pages. If the user has not been assigned a communications channel but its registration is still active, then any pages will arrive over the out-of-band signalling channel through a directed unicast message. If the registration update period has expired and there is no response, the only recourse for call alerts and pages would be via a general broadcast from all airborne platforms. This is generally less desirable.

[**Geolocation** - If the subscriber has a GPS function built in, the registration process can encapsulate this information which can be relayed to a database used to track asset location. We will assume that all the backbone nodes have GPS to be used in possible network topology calculations. Geolocation data must be protected from possible adversaries especially when operations are located behind enemy lines. Geolocation data would only be released when the associated packet can be encrypted. However, geolocation data is particularly useful to determine friendly force deployment to minimize fratricide.

There is another interesting aspect of geolocation information. Let us assume that both a user and his MCD are captured. The user can be coerced into revealing his private information (password); the MCD can then access the out-of-band signalling channel. However, if positioning is part of the initial signalling, then location information can be made part of the authentication process to sort out if the coerced user is in known enemy-controlled territory or in an area known to be OK to receive messages from. In the former case, these signalling messages can be handled differently. One could conceivably assign a specific demodulator with limited resources and redirect the communications to a server with disinformation.]

If GPS is jammed, the subscribers will lose the ability to encapsulate position with their registration messages. However, we will assume that airborne nodes have a backup positioning capability (inertial references) that can also be used to compute the node locations; similarly for the mobile support node. This means that there is always a way to compute backbone topologies without requiring transceiver distances computations. However, if synchronization requires this, then these computed distances could aid in the topology determination.

We will not assume that the aircraft maintains a database of accurate subscriber position location relative to the aircraft. This would entail an additional computation whose value is uncertain. Some systems such as MSS (Mobile Satellite Systems) do this computation in order to determine future beam coverage, but given that the flight pattern of a UAV is so uncertain, this would require too frequent updates to make it worthwhile. Note beam changes in multibeam subscriber systems can be fast and unpredictable and it is better for the subscriber to make measurements and adapt dynamically rather than for the aircraft to predict ahead. This will be expanded on in the next subsection.

Reregistration

Local reregistration occurs when a mobile moves between demodulators within a group, between groups, and finally between airborne platforms. The priority levels should not depend on which beam and platform to which the user is connected. However, this does not mean that the user in an active session will automatically be granted new uplink assets following beam handover since he may be contending with higher priority users on a heavily used channel.

During a session, the subscriber's assigned data channel may need to be changed as a result of relative movement of the aircraft with respect to the untethered subscribers. This is part of the handover process. Let us enumerate the possibilities with respect to uplink demodulator and downlink modulator assignments. On each airborne platform there are a large number of subscriber demodulators and a smaller number of subscriber modulators. Furthermore, a multiple subscriber beam array is used, then each beam is assigned a number of demodulators and one or more modulators. (One is the preferred number since it simplifies the downlink asset organization.) A beam "group" is an assembly of dedicated demodulators and modulator with a logical associated out-of-band signalling modem. Even when there is only one subscriber beam per aircraft, one may want to create separately managed groups to handle the needs of a specific organization that has some special security problems (i.e., intelligence). While this latter partitioning may be justified on operational grounds, it is inherently an inefficient assignment policy and should be avoided wherever possible.

The following situations reflect the user's subscriber channel assignments being dynamically changed:

1. Movement of a subscriber channel assignment within a group. This can result from a user being bumped from a higher priority demodulator with a specific QoS to a lower priority demodulator with a lesser QoS. This sort of change should be completely transparent to the BB router while the subscriber router needs updating. Moreover, it does not require any reregistration.

2. Movement of a subscriber channel assignment between groups on the same aircraft. If we assume that the subscriber channel controller handles all beams on the aircraft, this situation again should be invisible to the backbone router. However, there needs be a handover scheme between different groups and then we need to know how the aircraft's channel controller is implemented. If the channel controller handles all the groups, then the changes are also invisible to the backbone router and reregistration is not needed.

3. Movement of a subscriber channel from one aircraft to another. In this situation, the end-to-end routing must involve updates of tables in all the backbone routers. Before the first channel is deallocated, one must confirm that capacity on the second node is available. This is performed through signalling between the two involved subscriber channel controllers. During the transition process, the incoming and outgoing packets should be inhibited (or dropped in the case of isochronous data).

Before going into more detail on the registration process, it is useful to outline a possible scenario for beam handover between platforms or for beam-to-beam transfer on airborne platforms supporting multiple beams.

A number of operational assumptions needs to be made for tutorial purposes and the following scenario is only one of a number of possibilities. It is assumed that each beam has associated with it some "well known" channels to provide downlink and uplink synchronization and signalling. Let us assume that the user is active. Let us assume that the user is the initiator of switching between coverage beams. The handover process may be triggered by a received downlink average E_b/N_o falling below a certain threshold. This event triggers a search for alternative well-known channels (for other beams) on which sync is performed and an average E_b/N_o is determined. If it exceeds the original by a specified amount, the MCD may request, via signalling on the original beam, a request to change link affiliations to an alternate beam. If the beam were on the same airborne platform, the subscriber channel controller would complete the transaction. However, if the beam were on another platform, then the situation is slightly more complex.

In the registration scenario just covered, the authenticated user's access privileges were transmitted to all airborne subscriber systems. These privileges are the basis on which the signalling channel control can determine if there are sufficient uplink assets. Assuming that the user can be accommodated, the signalling controller on the new channel would route a "request grant" message to the user via the cross-link to the original platform's signalling message controller and hence to the user. This "request grant" would contain the identifiers of the demodulator and modulator that the user would now operate over. If the airborne node has only recently arrived on station and has no information on the user privileges, then this node must interrogate the authentication server to capture the privileges. With these privileges available, the subscriber channel controller can determine if there are sufficient uplink assets.

Now let us consider the situation where the user is passively listening but had earlier registered with one beam on a specific airborne platform. Assume that the MCD has automatically determined that an alternate beam provides better communications capability. If directed (other than general broadcast) paging and call alerts are to be routed correctly, then the user needs to take an explicit action to register on a "better" beam. This reregistration would be similar to a new registration. Observe this is slightly different from case 2, since even when the beams (original and new desired beam) are on the same aircraft, there is a need to inform the airborne platform which beam is being monitored. If the paging or call alerts are platform-specific rather than beam- and platform-specific, then the MCD needs to take an explicit interaction when a "better" platform comes into view.

Deregistration

An active registration implies some active database fields on one of more nodes and servers on the mobile WI. If a user is leaving the tactical theater, then an explicit "leave area message" is generated to alert the foreign agent that the user is no longer reachable via the WI. This in turn generates a message to the user's home agent. Finally, all traces of active database fields are deleted. One may leave some of the database fields in the authentication server in case the user is again attached to the mobile WI. This could happen, for example, if a warfighter alternatively moves his data device from a mobile WI connection to a local LAN still in the theater.

All registrations should have a finite lifetime to prohibit registration tables from getting too large. The larger the size of the search tables, the longer the processing times and the memory requirements. The editing of tables when the entries lifetime exceeds a specified limit is an operationally specified policy.

2.2 Authentication and Security

The intent of this section is not to delve into all aspects of security but mainly those that provide support for the authentication services with respect to allocating uplink resources. Part of the authentication is part of the registration process outlined in the previous section. Clearly, controlling access to the uplinks is the first line of defense with respect to security. Similarly, all the inactive users should be able to listen to the downlink out-of-band signalling, call alerts, and small paging messages. However, one does not want to grant a nonauthenticated user access to the downlink that carries user data. Hence, one could observe that the out-of-band modem is a logical modem having uplink demodulation and downlink modulation capabilities that are separate from the entities carrying authenticated users' signalling and data. We will cover in more detail in the appropriate sections.

2.3 Subscriber (User) Signalling

As noted in the main document, signalling comes in two flavors: one branch evolving from telephony and the other from networking. Since we are blending telephony like signalling operations with traditional networking support operations, the term "signalling" is used to cover both cases.

2.3.1 Telephony Signalling Channels

Out-of-band Signalling

Before the user is assigned a channel (up and down), the signalling must be accomplished via the out-of-band channels. The out-of-band signalling channel has a spread spectrum key which is known to all users on start-up. Presumably this is either part of the data loaded into the MCD before deployment or an inserted smart card. This would be the most vulnerable channel to be exploited for a denial of service attack. Presumably, this MCD, if captured along with its owner, can be coerced into attempting to log onto the system. If there are some "cues" that this is a captured MCD, then one would either deny further services or assign the user to a special channel meant to distribute misinformation. The immediate goal is to prohibit a captured MCD from both tying up valuable communication resources and giving unauthorized access to mission critical data. The out-of-band channel is used only when the user does not have an active session established.

The out-of-band signalling is limited to exchanges between the user and the airborne subscriber subsystem. We initially exclude user-to-user signalling or user-to-other-network-element signalling on the out-of-band channel on the basis of network security issues.

In-band Signalling

Once the channel is assigned, then the user is assigned to a channel and all signalling traffic is sent on the same channel as the data. In this architectural option, the uplink channel is an assigned and shared airborne demodulator while the downlink channel is either an assigned or shared time slot on the downlink broadcast. Both user data and signalling are intermixed on these assets. This is referred to as in-band signalling (IB). The MCD must, however, be able to separate data packets from signalling packets and pass them up the respective stacks. There is a need to have a thread (process) handling signalling separate from that handling data streams.

In the option #1 scheme, a user is assigned to operate on one of the demodulators at a time. Each of the demodulators is "isolated" from each other in terms of mutual interference by assignment of different SS patterns. After a user is assigned a shared use of a demodulator, the user has controlled access to the uplink and implicitly given access to cross-links and downlinks depending on what is allowed on the uplink. Once a user is assigned to a demodulator, all the user's signalling is performed on this channel with normal datagram messages. These messages use the same uplink access techniques as the regular data, but many of these messages are trapped by the signalling message processor for interpretation.

Whenever the MCD is involved in an active session, it will be listening to the standard broadcast downlink (not the logical out-of-band signalling downlink).

2.3.2 Network Signalling

For networking support, we use the term signalling to represent all the communication mechanisms and messages excluding those directly associated with carrying user traffic. Section 6 will delve into considerably more detail on the melding of the telephony and network protocols. The differences will become clearer as we examine the network protocol stacks that describe pairwise interactions between peer subsystems. The protocol stacks to be used in the mobile WI have some similarities to those that one traditionally sees when looking at a traditional TCP/IP network but also retain some specific carryovers from cellular telephony considerations.

The network signalling would include protocols such as ARP, mobile IP routing, mobile multicast routing, etc. These will be expanded in greater detail as we discuss the individual IPv6 protocols.

2.4 Beam Handover Operations

We have already stated that because of the relative motion of the airborne platform's antenna beam footprints with respect to individual users, there must be a mechanism that automatically moves a user from communicating over one beam to a new communications association on another beam. Moreover, this transfer must be automated and result in the loss of no essential data. Beam handover is a characteristic of all mobile cellular systems but there is a considerable variation on the details of the handover process.

In terrestrial cellular systems, the basic layout is in a hexagonal grid with the base stations at the center of each cell. The user switches connection from one cell to another only because of the user's motion across cell boundaries. Normally, because there is continuous coverage over a large area, the decision when to switch is based on relative signal measurements associated with a few adjacent cells. The cell with the highest average received signal is the preferred cell.

In mobile satellite services, the cells are defined by the satellite's array antenna beam footprints. Because of the high speed of the satellite motion relative to the earth in comparison to the user's speed, the beam handover is driven by the cell footprints moving past a relatively immobile terrestrial user. It is not unusual to transition between cells every few minutes.

In the Warfighter Internet, beam handover is less predictable because one can have significant user motion and a backbone topology with significant dynamic changes in the relative node distances. Maintaining a relatively fixed backbone topology may not be the first priority of the airborne platform if there are other noncommunication missions (i.e., surveillance). Additionally, the entry and exit of nodes at different locations at random times is an added complication. There also may be times when there is a "gap" in coverage from all airborne platforms. At this point, it is sufficient to state that the request to change beams comes from the MCDs based on their downlink signal measurements on synchronization channels. The airborne platform allocates new channels, transmits a grant message, switches to the new allocation, and "frees up" the old allocation.

It is useful to consider some of the differences between maintaining connectivity between a WI user and a user connected to relatively fixed position assets such as a geostationary satellite that performs on-board processing. In the latter class of communications systems, a user terminal synchronizes first to the downlink. The terminal is put into a downlink acquisition mode which typically involves a course synchronization phase for eliminating time and frequency uncertainties. Time is the major uncertainty since frequency uncertainty is mostly Doppler uncertainty given reasonable MCD frequency sources. Once course synchronization is completed, the downlink fine synchronization is performed, and once completed, an uplink synchronization process is performed requiring feedback from the airborne platform. After the synchronization process is completed, the communications session can proceed and the normal communications tracking process maintains synchronization. Resynchronization is needed only if one loses track, which is a relatively infrequent event

The WI user has a more difficult overall synchronization problem which is similar to that encountered by the users of the mobile satellite systems—MSS (i.e., Iridium or ICO). As in the MSS systems, new synchronization events can occur as often as every few minutes. This is driven by the requirement that the MCD periodically search for alternate downlink beams (on the same or different airborne platforms) that may be more favorable for continuing an on-going communications session. Once an MCD has initiated a communications session on a particular airborne platform beam, it has to periodically check to see if there other beams on the same platform or on the neighboring platforms that would provide a better link. This means that one must interrupt communications on the current beam, test the relative signal levels from all the neighboring beam, record their levels, and after comparing these with the levels from the original beam, make a decision whether to initiate a handover request. Each new beam examined means a new synchronization event, and during this period the MCD is removed from its primary networking task on the current beam.

The periodicity on which an MCD needs to evaluate other beams depends on the position of the ground user relative to the positions of airborne platform cellular footprints and the relative vector speeds of all these entities. Clearly, if an airborne platform has only a single beam, the interval between comparisons can be considerably longer than the situation where the platform has a multibeam subscriber antenna and the platform is running over a tight racetrack loop.

3. On-board Subscriber Subsystems

One will observe that, as yet, there is no subsystem allocated to initial terminal synchronization. The initial terminal synchronization requires frequency and time synchronization so as to be able to initiate out-of-band signalling. We have not specified the synchronization technique, since we have yet to select a spread spectrum scheme. However, as a placeholder we will associate this with the out-of-band modem subsystem and leave any details to later. Figure 2 shows the subscriber system at the top level (minus the logical OoB system).

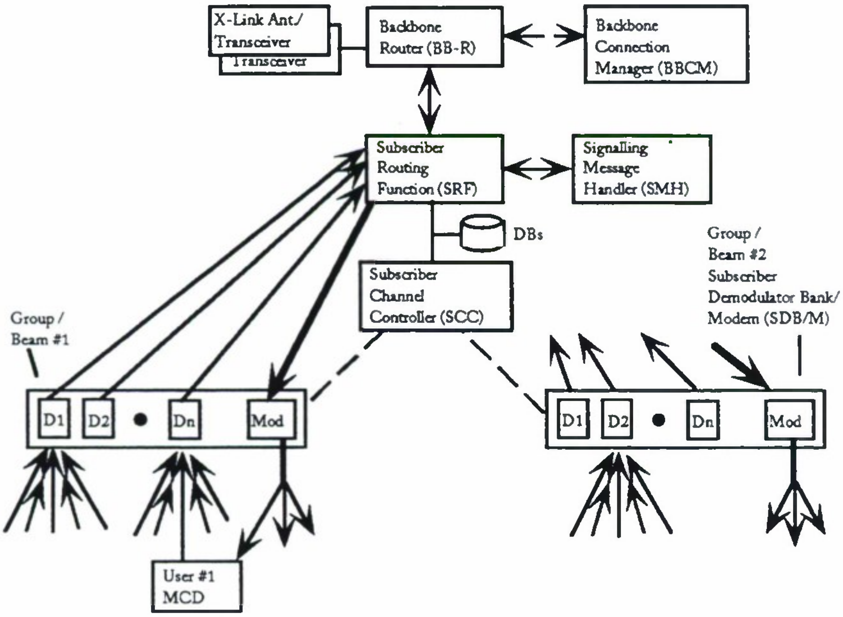

X-Link Ant/
Transceiver
Transceiver
Backbone Router (BB-R)
Backbone Connection Manager (BBCM)
Subscriber Routing Function (SRF)
Signalling Message Handler (SMH)
DBs
Subscriber Channel Controller (SCC)
Group / Beam #2 Subscriber Demodulator Bank/ Modem (SDB/M)
Group / Beam #1
D1
D2
Dn
Mod
D1
D2
Dn
Mod
User #1 MCD

Figure 2. Airborne subscriber system.

This figure shows a number of key features of the system. Unlike many cellular data systems, a user is not assigned exclusive use of an uplink channel for either data or voice communications. Rather each uplink channel is shared by multiple users via an assignment algorithm depending on the priority of the user, the precedence of the information, and the type of information. Access to an uplink channel is one of the key design issues in such a system and will be discussed shortly. The downlink is a single TDM stream with all the user's data traffic (and signalling) statistically multiplexed in an efficient fashion.

Two groups are shown which could represent two separate beams or perhaps two user communities that need to have exclusive control of their communication assets. The former is a more efficient way to increase overall system capacity whereas the latter configuration starts to negatively impact the benefits of statistical multiplexing. The architecture is configured to scale with added beams or separate user groups as required.

3.1 Out-of-band Signalling (Logical) Modem (OoBS)

We have already noted that the out-of-band signalling function required a logical modem function. This is the system that is easiest to penetrate in a security sense. All subscribers who have a recognized password for the MCD operation can send signalling packets on this channel and will receive responses (for this user and any other users participating in initial signalling). See Section 3.5.2 for one possible implementation.

3.2 Signalling Message Handler (SMH)

The signalling message processor is responsible for creating and analyzing messages associated with parts of the telephony signalling function. We also recognize that there are implicit signalling functions associated with the management protocols associated with IP operation that are embedded within the routers. [While the telephony-like signalling is shown as a separate functional unit, it probably could be implemented as a separate task within the physical Subscriber Router Function implementation. However, this telephony function is not a "normal" router type function and it is best to keep them in a separate specification.]

All the data from the uplinks and downlink channels pass via the subscriber router function. The signalling messages from the MCD are all directed to the signalling message handler. We would expect that until a validated request for data communications is acknowledged and resources are allowed, the signalling message handler is the only on-board entity that the MCD can directly address in terms of telephony signalling.

3.3 Development of a Subscriber Uplink Multiple Access Scheme

3.3.1 Overview

This section is a long one dealing with the issues of efficient uplink subscriber access both with respect to voice and data. As pointed out, the uplink is the main communications constraint for the entire Warfighter Internet and as such deserves the most careful consideration. The main focus will be on data because of the unpredictability of instantaneous traffic loading (bursty traffic at low duty cycles). However, concentrating of the provisioning of efficient data services should not preclude providing similarly efficient voice services without requiring a physically different and segmented design. Most of the following discussion will be concerned with providing efficient data services. For the strawman system, we will assume a very simple subscriber voice access system that is more circuit-oriented although it transitions to packet voice on the backbone. (See Section 3.3.6.2.)

The structure of these subsections is to look first at a top level at the issues associated with providing an efficient uplink multiple access scheme. This is done since uplink access is potentially one of the overall system bottlenecks. Efficient multiple access has a long history in network research. We will discuss one such system (Ethernet) since it is a well known scheme that has performed exceptionally well and has some exploitable common characteristics with the WI subscriber uplink. We will then transition to the cellular CDPD that has some similarities to Ethernet and, potentially, even more to an efficient realization of the WI. Finally, there are some aspects of the WI scheme that are not covered by either; however, the mobile cellular GSM will be seen as a rich source of system solutions that have applicability. As we synthesize a conceptual design for the mobile WI, we will abstract and blend elements from all these systems as appropriate.

This section also precedes those sections that describe the subscriber channel controller and the subscriber demodulator/modulator subsystems because the choice of an uplink access scheme implementation will influence those designs.

We should note that many multiple access schemes are contingent on a parameter, ß, which is the ratio of the propagation delay to the transmission time of an average packet. The usual division is between those environments where ß << 1 and those where ß > 1. Ethernet is an example of the former, and satellite communications of the latter. Similarly, for the latter, reservation schemes seem to be dominant, whereas for the former, some variant on a distributed access scheme is prevalent. Reservations imply a central scheduling process.

Let us clearly distinguish between those users making a “reservation” request because of an “anticipated” traffic need and those users that make a “reservation” request based on a queue of real traffic. At this point, let us consider an uplink data transfer rather than voice traffic. An “anticipated” traffic request would be, say, for a large file transfer on the uplink. One does not know the rate at which the file will be transferred, but a request may be for a dedicated fixed bandwidth channel (i.e., 50 kbps) for the duration of the transfer implying a fast transfer. However, there is no guarantee that the entity receiving this large transfer can accept the maximum rate at which it can be delivered, and thus the allocated 50 kbps would be underutilized. In this atypical example, the server was attached to an MCD and the client could be at another MCD or external to the WI.

Now consider the situation where the server is attached directly to the backbone or via a high data rate link to the backbone and the client is on the MCD. The client initiates the file transfer and must “reserve” its uplink and downlink assets. The question is how much uplink does one reserve? Clearly, there needs to be some initial searching of the server host file system to locate the file and then the transactions that start the transfer. The uplink capacity reserved should be small and the downlink capacity reserved larger. Again, the downlink capacity reserved should be no larger than the ability of the MCD to capture the data. The tendency is to request oversized uplink and downlink capacity allocations for inappropriately long times.

Now look at the situation where an uplink "fast reservation" can be made based on each user's individual queue of actual outgoing traffic. An uplink reservation can only ask for capacity up to the size of the current queue. There needs to be an uplink reservation channel that is used to convey this request to the base station but we have not yet specified how the logical uplink request channel is to be provided. We will make the assumption that the uplink is slotted into fixed size slots (for user data) corresponding to a fixed transfer unit duration. Since the uplink channel is shared by multiple users, the base station has to capture all these requests, prioritize the requests and then send down scheduling information to each of the MCDs so that they can release a specified amount of information in the proper slots without fear of contention. The "grant" messages to the individual users is sent on the downlink in another logical channel. This "grant" message not only includes a user ID but also a starting slot number and the maximum number of permitted sequential slots. Since the assignment is to be accomplished on an uplink slot-by-slot basis, it is clear that the amount of downlink slot management information can be significant. Furthermore, one should also not minimize the difficulty of the base station creating a new uplink schedule each and every time a new uplink request arrives.

Overall, this "fast reservation" model appears to have too many drawbacks and thus we do not consider it further. Its major benefit was that it promised to use the uplink channel more efficiently, since it only used the uplink data slots when there was actual data to transmit. Variations on fast reservation schemes are totally appropriate for satellite systems because of propagation delay problems. However, the WI internet delay problem is not severe and hence we have more flexibility in the choice of a multiple access scheme.

Therefore, we will examine schemes that do not require the MCD to ask for explicit uplink capacity for their transmissions. Each of the MCDs are free to transmit based on actual queues of data and subject to a well defined transmission policy and subject to a method of detecting collisions between simultaneous uplink accesses and terminating on detection. The detection mechanism and broadcast of this collision alert is one of the key design issues behind this scheme. Basically, the base station detects the collision and sends down collision alerts. However, the base station is not given the burden of scheduling the transmissions of all the uplink transmissions. This is a major simplification. However, we still need to enforce user priorities and message type precedence and some measure of QoS. This is another aspect in the design choices to be covered in more detail later.

It should be noted that the Aloha multiple access technique was a revolutionary multiple access radio-oriented scheme which in turn led to many variations which are widely deployed today. Its LAN offspring, Ethernet, is the best known realization of contention-based multiple access and is an appropriate starting point for our discussions on a WI appropriate radio design. This will then be followed by consideration of the CDPD (Cellular Digital Packet Data) system which is an important variation on random access radio systems as applied to a particular instance of a terrestrial cellular system. We will see that CDPD is an appropriate candidate from which to borrow good ideas and map them into the WI design.

There are other cellular-based systems from which we will borrow concepts. It will be pointed out that CDPD is not a stand-alone system but coexists with the cellular analog voice system—AMPS (Advanced Mobile Phone Service). AMPS provides auxiliary services to the CDPD user. In the digital cellular world, the European GSM (Global System for Mobile Communications) sets the standard for many of the system architecture concepts and elements being used in later cellular voice systems. In particular, with telephony systems, unlike data networks, the user employs "signalling" methods to allocate communications resources before any class of communications can begin.

In the mobile WI, there is also a telephony aspect to reserving communications resources on the subscriber subsystem and thus one needs to look at any relevant commonality with the solutions offered by GSM. [The Mobile WI does differ in that it does not reserve sole access to a communications resource but rather a "right" to share this resource with other users.] We will see that the initial acquisition process involves user authentication and assignment to a specific WI demodulator has analogies to comparable function in the GSM system.

GSM also provides the definition of a complete infrastructure support for mobility which mobile IP emulated. Therefore, there are additional reasons for looking at specific GSM system designs. One does not want to reinvent functionality for WI which has been successfully analyzed and implemented first in GSM.

3.3.2 Uplink Access Techniques—Efficiency versus Implementation Tradeoffs

In the consideration of efficiency of data transmissions, it is very easy to justify using packet data communications versus circuit-oriented data. This is a result of two issues: 1) data generation at either end of a session is usually very bursty with very low duty cycles, and 2) the data transmission are usually very asymmetrical in average data rates. Therefore it is not unusual for packet communication techniques to provide an efficiency improvement of a factor of 10 to 1000 over circuit-oriented data communications. However, once one selects packet transmission as being the mode, the exact form of packet communications becomes the next issue. In the overview session, the claim was made with respect to the WI that the complexity of the "fast reservation" scheme was such that simpler contention-based schemes would be more desirable.

There is no disagreement that "fast reservation" schemes can often be made more efficient than contention resolution based schemes and this factor could approach a factor of 2. A factor of 2, while not a factor of 10 to 1000, is still something worth considering as long as it can be realized in actual implementations. Let us consider this aspect with respect to WI.

In the WI, we have postulated providing approximately 32 uplink demodulators and a single downlink modulator for every group (or beam). The assumption made was that one could build and replicate demodulators at little cost and added power. Moreover, the assumption was made that the uplink burst rates would be about 100 kbps and the downlink rate about 1 Mbps. The following is only meant to be a back of the envelope type of argument. Assume that an arbitrary uplink access scheme can fill the channel with an efficiency e_i. The aggregate uplink data rate that can be supported is then of the order of (32 * 100 kbps * e_i). The same community shares a downlink that is only 1 Mbps. For the purposes of the argument, we will assume we can have perfect scheduling on the downlink and the efficiency of utilization can be made close to 1.

Now if the data transfer between mobile users is roughly symmetrical with respect to aggregate uplinks and the downlink, we can see that there is a factor of approximately 3.2 * e_i more uplink capacity than downlink. However, in most client-server type of applications, the client-generated traffic may be a factor of 10 (or more) less than the traffic returned by the server. One would expect that in the WI the majority of servers are connected to the backbone mainly via the terrestrial gateway. Thus, one can assume an average asymmetry reduction factor, A, in favor of the mobile users. Thus, the downlink saturation point would limit how much information would be transferred on the aggregated uplink as follows:

(32 * 100 kbps * A * e_i) ~ 1 Mbps or
e_i ~ 1/(3A)

Thus, one can saturate the downlink channel with a fairly modest average loading of uplink channels. The key point is that one wants to assign groups of users to each channel with the goal to keep that uplink channel's utilization factor modest. Roughly, this means that if one has a class of users who are operating as typical clients with a modest rate of server requests, one could assign a large number of such users to a single channel with the expectation that even peaks in utilization may never exceed say 10%. This is certainly the experience in Ethernet occupancy measurements. We will discuss the user-to-channel assignments later in the section on the Subscriber Channel Controller.

If one believes that the uplink channel will, in a statistical sense, never be filled more than 1/3A, the question arises why not build fewer demodulators? Note that we use the "reserve capacity" to mitigate collisions which the potentially more efficient "fast reservation" avoids. However, there is a need to retransmit uplink segments (or blocks) in both cases due to uncorrectable errors due to propagation errors (fades). The so-called "reserve capacity" in either case can be used productively to handle these propagation degraded packets. The "over-provisioning" of demodulators is justified.

It is important to mention one other assumed system choice. In the main scheme proposed, a user is assigned a single uplink channel, and the channel assignment is done on the basis of the user's precedence level and perhaps the priority of the anticipated message. Traffic such as voice could be assigned to uplink channels specially configured to handle voice traffic. Each uplink channel is assigned a unique spread spectrum key which is shared in common by all the users sharing that particular channel. The intent of these unique keys is to minimize the mutual interference between users in the different uplink channels. It is known that it is more efficient in terms of overall channel efficiency if each user is able to access not one channel but all channels. This applies whether one is using reservation schemes or contention access schemes. However, this would mean that the user would need to change the key whenever the message segment is to be transmitted on a different uplink access. If the keys were produced independently, then loading them in the spread spectrum pattern generator on a transmission segment-by-segment basis could stress the key handling capability of the MCD. However, if one could easily generate quasi-orthogonal or orthogonal spread spectrum streams from a single key through some simple operation, then the task of switching rapidly from uplink channel to uplink channel is simplified and the user can operate on multiple channels on a fast switching basis. We will refer to this choice as the single user–multiple uplink channel assignment scheme. Needless to say, there are many variations on this basic scheme which should be evaluated not only for efficiency but also for implementation difficulty. However, we will still concentrate on the single user to single uplink channel assignments for the purpose of fleshing out some of the fundamental descriptors of the overall system.

3.3.3 Ethernet as a Partial Model

Ethernet's multiple access scheme (CSMA/CD—channel sense multiple access with collision detect) is a simple but very effective way of using a common broadcast channel where all the users can sense if the channel is busy and, more importantly, sense to see if their transmission is colliding with another user's transmission. If the latter, each of the colliding transmitters terminate their transmission and adopt a retry policy. Early termination upon sensing a collision is what makes Ethernet relatively efficient. The goal of the retry policy is to minimize subsequent collisions between the rescheduled packet transmissions. Whenever the channel is not busy, other users with packets to send can try to transmit and they may or may not succeed depending on whether or not other users attempt to transmit at approximately the same time. While Ethernet and the 802.3 standard have some minor differences, we will not distinguish them here.

Some of the key features of the Ethernet scheme are the following:

1. There is no global central scheduler who tries to efficiently field requests for data transmission time from users. The scheduling is really a distributed multiple access scheme with certain fixed transmission access rules.

2. There is no concept of a data flow specification required, a user only transmits as real data is available to be transmitted. The data is a real data queue not what the user believes the future data flow will be. Predicting actual data flows is virtually impossible.

3. There is no restriction on the amount of information wants to transmit within a finite period relative to any other user.

4. There is no call admission reject policy as in circuit- or most reservation-based schemes. The retry policy is the closest analog to call admission and it is strictly a mathematical expression (an exponential backoff retry algorithm).

5. Every user has the same rights to accessing the channel as another. Following the end of a transmitted packet, a new user has the same rights as any other in gaining access to the channel.

6. The fact that one has transmitted one or a series of packets with no contention does not mean that the next packet will not be contended. Collisions will lead to delay variability.

7. At some point in increasing traffic load, the actual delivered load will start to diminish because more of the capacity is wasted on collisions.

8. All transmission decision making is handled entirely by the user's network interface card which is always sensing the state of the channel.

9. Every other user's transmissions are sensed. There is no hidden terminal problem.

10. There is no concept of a cyclic frame that a transmission must adhere to. In the Ethernet standard there is no slotting (like slotted Aloha). It was felt that early termination of colliding transmissions was better given the short propagation distances.

11. All traffic types are treated equally.

12. There is no way to guarantee a specified quality of service for any particular traffic stream.

13. The physical size of the Ethernet network is conditioned by maximum delay times across the network.

There are a number of other LAN access schemes that have been designed; some promising higher transmission efficiencies, but they have never received the acceptance of Ethernet and its higher speed derivatives. IBM's token ring was the only serious contender and a variant FDDI has had limited success, but all have been eclipsed by faster versions of Ethernet-like technologies.

Now let us look at the WI subscriber access problem and see how its environment compares to that of the Ethernet LAN. It should be noted that in terms of ß, the mobile WI is not too dissimilar from that of Ethernet, even though one is LAN and the other is a radio channel. The propagation delay is over two orders of magnitude greater than Ethernet, but the uplink burst rate is also more than two orders of magnitude smaller (making the packet transmission times over two orders of magnitude longer). This means that ß is about the same as Ethernet. However, there are some still important differences which follow:

1. Multiple access is only needed for the uplink; the downlink is a single stream managed by a single entity. The downlink has a considerably higher data rate and is not as constrained a resource and thus can carry some measure of coordination information.

2. Only the airborne platform listens to all the user uplinks; the users are not capable of listening to other user's uplinks and hence have no direct knowledge of their transmission colliding with another user. All the other users are thus "hidden" transmitters and thus a user cannot "sense" the channel either as being a clear channel or a channel that is being used by a user (or multiple simultaneous colliding users).

3. The users on the Mobile WI wish to be able to enforce some order of priority.

4. The Mobile WI must support isochronous traffic with acceptable performance. While voice over a packet network has been demonstrated a number of times, it is not a mainstream technology.

5. On a LAN, the error rates, due to noise or interference, are very small. Retransmissions are needed mainly to account for collisions. On WI, channel errors due to adverse propagation conditions (i.e., due to multipath) are considerably higher. These channel effects will result in increased retransmissions.

Ethernet is a decentralized scheme for managing multiple access. Each user's network interface observes the channel and transmits when it has data to transmit subject to a specific access algorithm. In the WI case, only the base station is in a position to make judgments on the state of the uplink channels and act appropriately. However, before discussing the WI problem in more detail, it helps to see how the CDPD system handles their uplink accesses.

3.3.4 Relevant Attributes of the CDPD System

The CDPD (Cellular Digital Packet Data) system is, as its name implies, a packet-oriented data service overlaid on a voice system (in most cases the cellular voice services provided by AMPS). AMPS (Advanced Mobile Phone System) is a first generation analog mobile voice cellular system that is ill suited to carrying data. AMPS is an FDMA system with each channel being 30 kHz of bandwidth (in each direction). In the initial allocation, each provider had about 312 channels which have to be apportioned over a finite number of cells (a cluster). Each cluster of cells will contain all the 312 channels and this cluster is repeated spatially to provide frequency reuse. The point is to have cells which contain the same set of frequencies spatially separated so as to avoid mutual interference. CDPD is a well-designed cellular wireless-based packet network that deserves careful attention since it overcomes many of the problems to be faced when designing WI. If some of their systems engineering decisions fit WI, then one must give careful consideration to them. One need not reinvent proven designs.

In voice operation on AMPS, a new call will be assigned a pair of channels from the pool of inactive channels. However, usually even at busy times there are free channels that are not needed for voice communications. These free channels can be used to support CDPD. Unlike AMPS, where a single user occupies a pair of frequencies, in CDPD a pair of frequencies can be shared by many users to significantly increase efficiency of utilization. This is the same concept that is driving data handling on the WI.

Since CDPD came along after significant deployment of AMPS, it is designed to require absolutely no change to AMPS. Therefore, when a voice request results in the assignment of a channel being used by CDPD, the latter system must terminate operations on that channel and find another free channel. The burden is placed entirely on CDPD. Still it should be clear that CDPD is not a stand-alone system but is an auxiliary service superimposed on a complete AMPS system.

Thus, it should become clear that CDPD is not a COTS solution for WI. However, this is not meant to imply that some of its system design elements are not appropriate for WI. In the following discussion on CDPD, key details will be presented, many of which can be adapted for WI and others that are inappropriate and illustrate the important differences between the design goals behind CDPD and WI. It should also be noted that we will concentrate on the subscriber subsystem since this is the part of the overall CDPD that is of most interest with respect to WI. The important interface is the "airlink" or interface A.

Earlier, we had noted the importance of the value of parameter ß (the ratio of the propagation delay to the transmission time of an average packet) being approximately the same between Ethernet and the WI. The value of ß for CDPD is about two orders of magnitude smaller than either Ethernet or WI since the cell radius (propagation delay) is about a factor of 10 less than WI and the burst rate is about 10 less. This means that applying some channel sense and collision resolution scheme from CDPD is extremely easy to justify. However, this does not imply that the techniques used in CDPD do not provide a quantitative improvement in WI but one cannot expect the same measure of improvement.

CDPD employs a DSMA/CD (Digital Sense Multiple Access with Collision Detection) analogous to a number of multiple access schemes that broadcast busy tones in the presence of collisions. The goal is for the base station to recognize collisions of uplink transmission as soon as possible and to alert the users so that they will terminate their transmissions and adopt a fixed retry policy that is intended to deconflict subsequent retransmissions. The transmissions are only started on microslot boundaries (related to round trip delay times and processing times). The microslot interval is 3.125 msec. The busy/idle indicator is transmitted on the forward channel every 3.125 msec corresponding to a subslot-by-subslot report on the status of the reverse channel occupancy.

The major categories of applications for CDPD are distributed systems with bursty data transfers and handheld interactive computing. The application is initiated, registered, and data sent as needed. Inactivity during a session is not penalized since the capacity can be used by others.

One of the goals of the CDPD network specification is to use the existing data network infrastructure as a key part of the overall CDPD system. Figure 3 shows the reference architecture for CDPD with a shading meant to illustrate its similarity with WI. We will make cogent comparisons of CDPD with WI and we will point out differences with WI by enclosing the comments in [..].

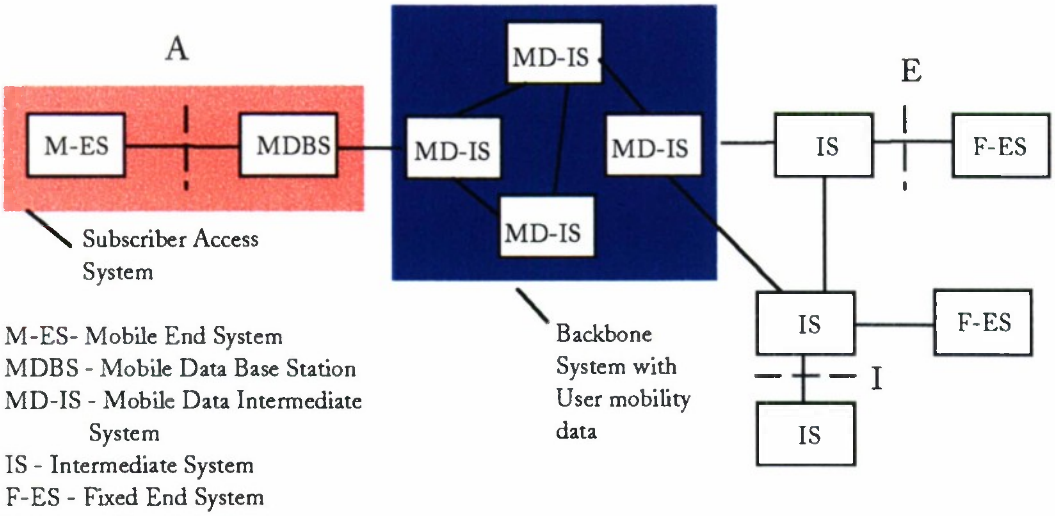


Figure 3. CDPD network reference model (expanded).

A simple description of the main subsystems is given below. To make the analogy more complete one should associate the term MD-IS with the WI router. Similarly, the MDBS should be compared to the multiunit airborne subscriber subsystem. The IS would correspond to routers outside of WI.

Mobile End Systems (M-ES) are the units by which the users get access to the CDPD network. Currently, the M-ES implementations and their associated applications exclusively use TCP/IP protocols. These are only data devices whereas the MCD in the WI handles both voice and data.

The **Mobile Data Base Station** (MDBS) performs the MAC and Layer 2 (Link) protocols for a set of radio channels associated with a cell even if that cell is sectorized.

The **Mobile Data Intermediate System** (MD-IS) performs routing functions based on additional knowledge of the current location of the M-ES. Mobility information is exchanged between MD-ISs via the **mobile network location protocol** (MNLP). In the specification, communications between MD-ISs require the CLNP which is the ISO connectionless network protocol, but because of the rapid growth of TCP/IP, the specification allows the use of either CLNP or TCP/IP.

Intermediate Systems (IS) are essentially routers transferring packets between various nodes. These routers, unlike the MD-ISs, are unaware of the mobility aspects of the end users. In CDPD, the ISs route either CLNP or IP packets as needed. If passing packets between MS-ISs with intermediate ISs, then the datagrams are CLNP.

Fixed-End Systems (F-ES) is a generic term applied to nonmobile end systems. Typically, these could be hosts connected via a land line data network. The F-ES usually exchange IP packets with the nearest IS.

Each M-ES is identified by a Network Entity Identifier (NEI). At present, the NEI will be an IP address. The WI MCD is identified by an IP address.

We also have an approximate CDPD analogy for a WI mobile support base station and this is shown in Figure 4.

It should be mentioned that the CDPD billing mechanism is based on the volume of traffic sent and not on the connection time. [While one does not need this in WI, it would be a useful function since it could give quantitative measures of the actual utilization of the system.]

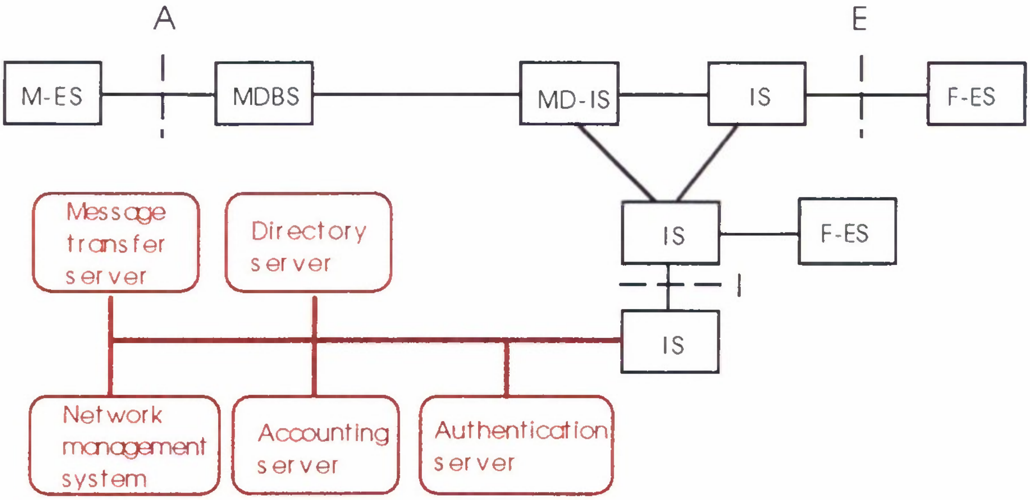


Figure 4. Network servers for CDPD.

Figure 5 shows only the airlink protocol profile and it includes the M-ES, MDBS and MD-IS interactions.

This protocol stack was designed when there was a major drive to implement the OSI stacks. Since then, the TCP/IP are becoming the de facto choice and hence there is a mix of both. [In WI, only the TCP/IP protocols are needed and hence there is some complexity reduction. However, the WI subscriber subsystem is more complex than MDBS and hence the protocol layer up and including IP are resident in the WI subscriber subsystem.]

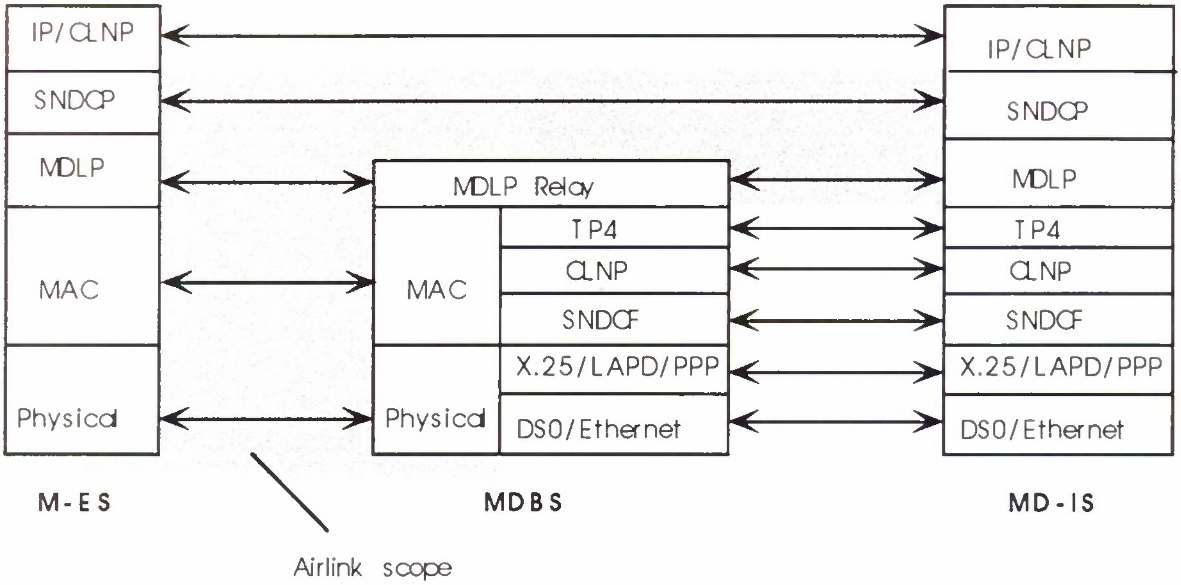


Figure 5. Airlink protocol profile.

In the above figure, the MAC (Media Access Layer) refers to both the uplink (reverse link) and downlink (forward). In both systems, only the uplink really requires a special consideration, the downlink merely does multiplexing. However, it is easier to retain a common term.

It is useful to compare the CDPD protocol stack with the data transfer entities across protocol boundaries. It is seen looking down the stack that an IP packet undergoes a series of transformations before being released. All of the transformations have a special purpose and we will only comment on them as potentially applicable to WI. Looking down the stack, the introduction of a shading (or color) implies a change in bit meaning from the transformation above.

Figure 6 shows what happens to a network packet as it flows down the stack and is ultimately transmitted. The receive process is also implied. Clearly, it is a complex process. Let us start with an IP packet of unspecified length; the only certainty is that there will be a large variation in the IP packet sizes within WI. The IP packet is handed down to the SNDCP layer.

The first transformation is header compression. [This could be implemented in WI to reduce header sizes, particularly for IPv6 headers with options.] The next transformation is the compression of the IP data portion. [Given that the user data will be encrypted, compression of the WI data may not be very effective and hence this transformation would probably be omitted.] The next step is segmentation. It should be pointed out that the segmentation was performed in the SNDCP layer so that the units matched the capabilities of the data link protocol frame size. The segment is not related to the size of either the uplink or downlink transmission blocks. Finally, the data part of the segment is encrypted before passing to the link layer. [In WI, there is no need to encrypt at this level and this can be omitted as well. Given the simplification of the transformation in the SNDCP layer, we will have to assess whether to keep this layer or to include some of its responsibilities in the link layer.]

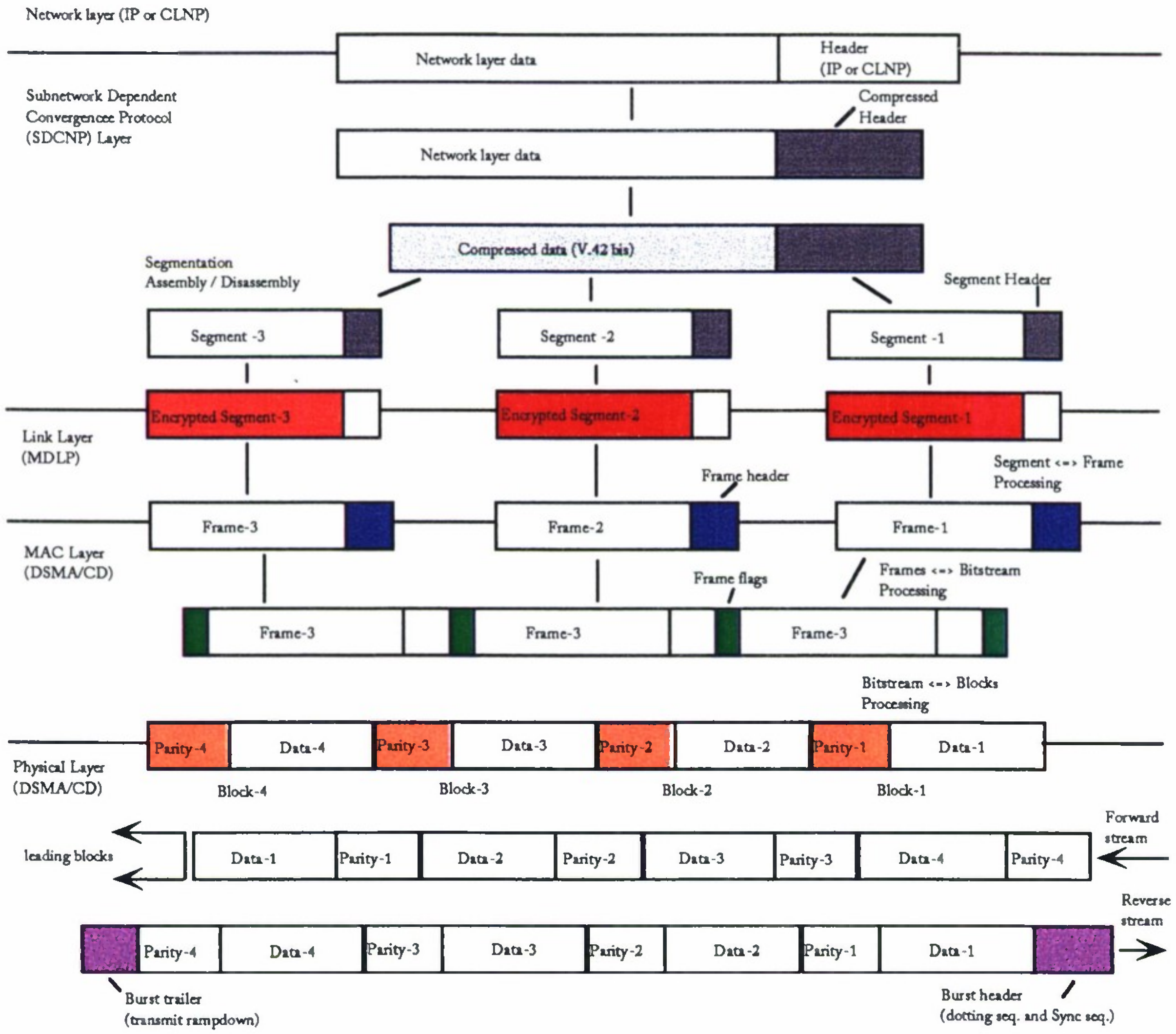


Figure 6. Data unit flow across airlink protocol stack.

In any case, frames with frame headers are what are generated by the link layer and passed to the MAC layer. There is information in the frame header to associate a sequence of frames with a particular packet so one can rebuild complete packets as needed.

In this figure, the frame is not the transmission unit but rather a collection of frames. Of course, because the channel is highly lossy, one should provide some forward error detection and correction (FEC). In this case, a Reed-Solomon block code is used in the MAC layer and we see the concatenated frames being divided into coded blocks. [In WI, we anticipate the use of FEC, and Reed-Solomon will certainly be one of the possibilities.]

The bottom layer is the physical layer and here is where there can be significant differences with WI. The physical layer deals with frequency selection, modulation and demodulation techniques, synchronization, etc. There will be little similarity. However, let us look at some issues that arise in both systems.

Let us first look at the uplink (reverse channel). Both are multiple access by uncoordinated users. Each user should produce a synchronization header for the base station to acquire before it decodes chips (or symbols). There is also a trailer to alert the base station to break track and get ready to acquire another bitstream block from either the same or a different user.

The synchronization problem for the downlink (forward link) is simpler for both CDPD and WI. The uplink access (the reverse link) was contended for but the downlink (forward link) avoids contention altogether since the base station handles the scheduling of all the downlink traffic. Once the downlink synchronization is performed, continual tracking is the rule. Since there is no power deficiency on the airborne platform, we can assume that synchronization-capable patterns are sent whenever real data is not being sent. The major difference between CDPD and WI is the downlink rate is about an order of magnitude larger than the uplink burst rate.

At this point, we could go into more detail on the services provided by each layer in CDPD. However, it is more fruitful to go directly to a synthesized WI scheme having certain CDPD attributes and state what is needed to support the tactical WI needs. We will only refer to CDPD where the services needs match well and the work done in CDPD can be applied.

3.3.5 Relevant Aspects of the GSM System

Mobile telephony-oriented voice networks have had to concentrate on the establishment and maintenance of channels in the face of propagation anomalies and had to solve the network-specific channel conditions in addition to all the “higher” level network conditions. We are using GSM as a logical model insofar as it applies. We use GSM as a model because it was essentially the first major digital cellular system to be planned at a system level and subsequently deployed; its basic architecture has been adopted by many of the following systems, including mobile satellite systems. Of particular interest to WI is the network layer. GSM has created a “network” layer with a least 4 different functions (RR, MM, CM, and SS). We show this in Figure 7. This figure shows a comparison between the WI subsystems and their GSM equivalents.

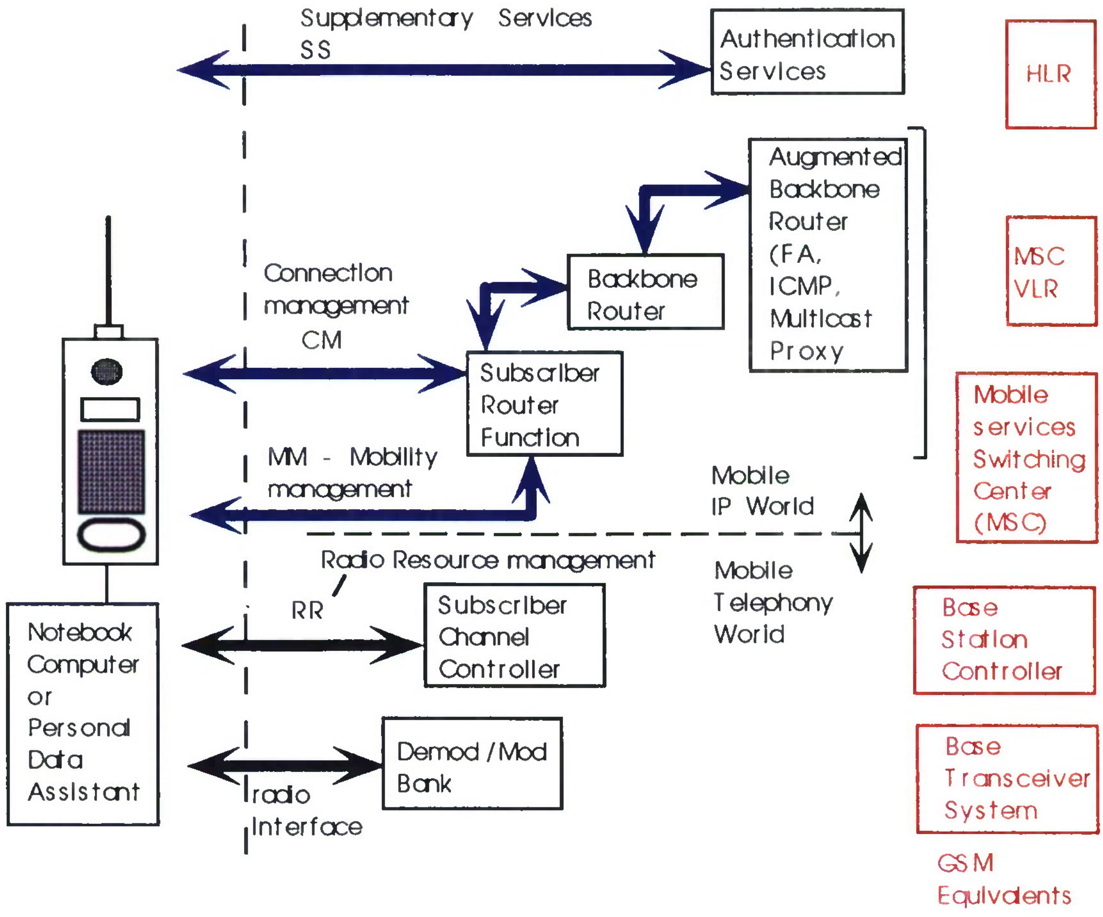


Figure 7. WI carryover from GSM cellular concepts.

It is important to recognize that we have limited similarities to the GSM architecture. GSM is a circuit-switched architecture; the mobile WI is a packet-oriented architecture that has some special problems associated with the dynamic assignment and deletion of links. Consequently, we have to merge these two separate networking developments into a single integrated scheme. To first order, we will need to map the MM, CM, and SS functions from the telephony world to the IP world. Let us first review the main functions of these sublayers as defined in GSM.

RR—Radio Resource Management Sublayer—The functions of this sublayer are connected to the physical layer operation. This sublayer is responsible for the management of frequency spectrum, mitigating against the changing radio channel, channel assignment and release, power-level control, time alignment, and cell handover. Note that fixed IP networks don't address these issues and hence this is clearly a cellular telephony-like function, and in our case a subnet-specific implementation issue. In the WI, the channel assignments are only the uplink and downlink subscriber channels.

MM—Mobility Management Sublayer —The function of this sublayer is to support user mobility, registration, and management of mobility data. While not shown explicitly, MM also uses authentication services to check out the user and determine the type of services allowed. The majority of functions belong in the mobile IP world.

CM—Connection Management Sublayer—This sublayer in GSM is responsible for all the functions necessary for end-to-end call control and management. Inasmuch as the WI IP protocols handle routing, the CM function will be absorbed into IP.

SS—Supplementary Services—These services are typically additional telephony features such as call blocking and call forwarding and are built on CM. We include it here to indicate that we may want to limit or restrict voice calls by different groups of users to specific areas. This would require some interaction of CM with an augmented Authentication Service.

Of these four layers, the last three will be embedded in IP considerations. This will be reflected in the candidate protocol suite that follows. To help clarify assignments, we show the GSM subsystems to the right side of the figure. The left side of the previous figure shows the entity pairing for much of the sublayer interactions. We also denote a conceptual boundary between the cellular telephony and IP worlds.

IP networks are just beginning to come to grips with the effects of mobility. However, the planned mobility is more in the vein of a mobile user becoming attached to a "foreign" network and becoming registered as a member of that network. There is little concern with the details of the physical connection since that is a subnet specific implementation issue and not part of the traditional IP network consideration. This is a strength of the IP internet concept: let the subnet handle the channel dynamics. It should be noted that mobility was not part of the CDPD design; it came along as part of the AMPS infrastructure.

At this point we can start a roughing out of a representative protocol stack so that we can identify subsystem entities that need to exchange information in order to create end-to-end paths for user transactions. In Figure 8 we will first show the candidate architecture laid out in a linear fashion and then show a representative layering for the signalling. Note that when we discuss the MCD that there are different signalling, data, and voice/video stacks. Since signalling is the most complex of the three, we initially want to concentrate on it.

The following figure shows only the basic protocols. It does not show the mobility support or multicast protocols (when either is needed for signalling). These are implicitly included but not shown. Many of these are essentially signalling applications that UDP needs (which is also not shown). Also, in the main text we noted that the uplink and downlinks are considerably different, so that their respective protocol stack would be different. We only show the uplink; the downlink would probably have a simpler alternative to S-MAC.

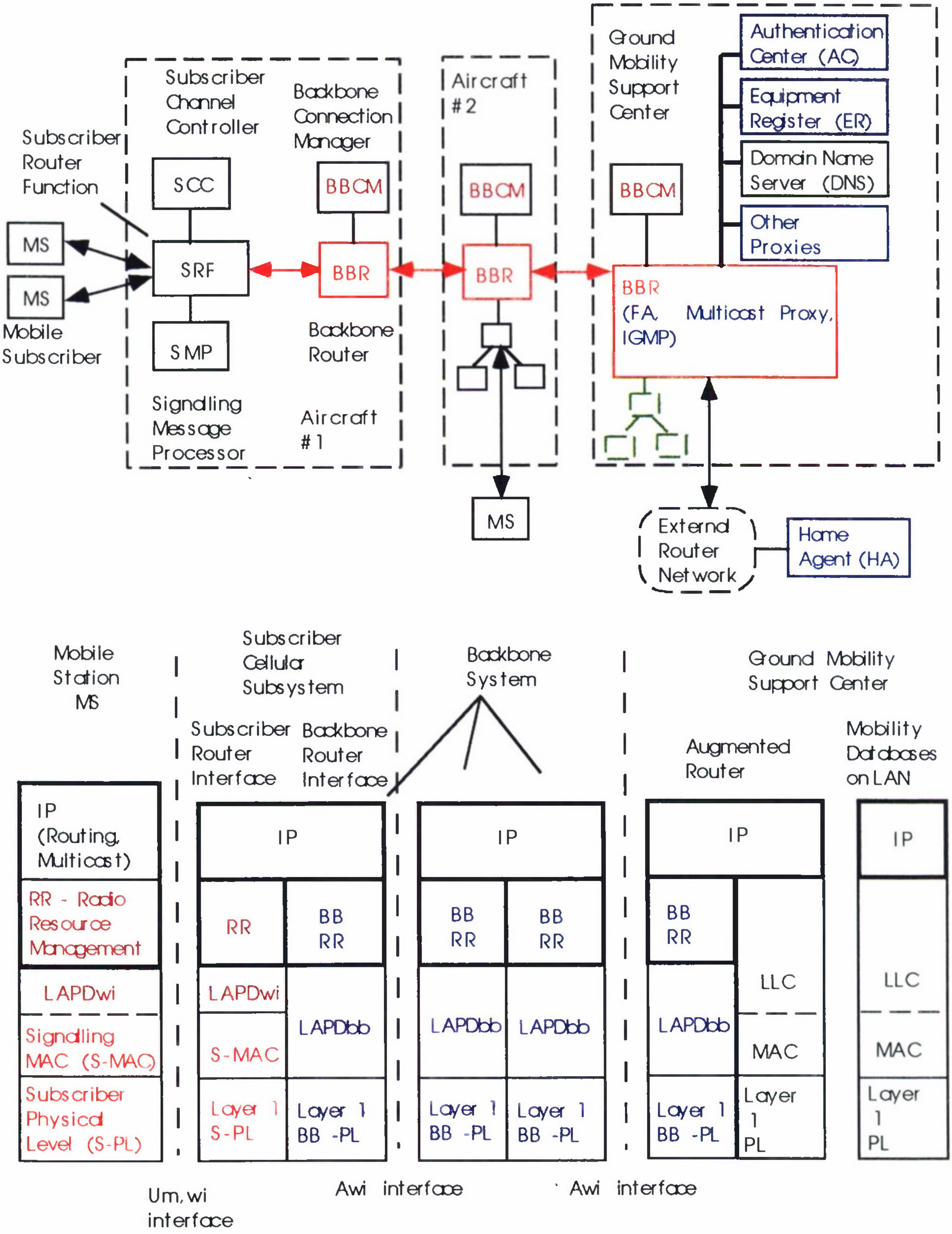


Figure 8. Candidate mapping of signalling protocols onto WI Architecture.

We have already discussed the need to produce a WI-unique RR sublayer for the subscriber links. As noted, this is subnet-specific and, like all IP subnet contracts, promises to deliver a connectionless service to the IP upper sublayer. Similarly, since the backbone is dynamic, we need another RR equivalent (called the backbone-RR) which manages the dynamic connectivity between the airborne platforms and those ground sites connected by backbone links. The execution of the backbone-RR would be in the backbone connection manager. Since the backbone links are point-to-point links, there is no need for a MAC sublayer. One would expect that the logical link level protocols (denoted by LAPDwi and LAPDbb) would have much commonality, since both links ultimately have to deal with channels exhibiting strong multipath. However, there is no need to make them identical, and thus we have maintained distinct notations.

Note that within the airborne node, we did not separate the subscriber router function from the backbone router in terms of a protocol stack. In the actual implementation, one would have to do this but this is a detail that we can hide at this time.

3.3.6 Synthesis of a WI Scheme

In this subsection, we are not going to provide the detailed parameters value for either the uplink or downlink since we need to provide a more comprehensive analysis. However, it is possible to outline a framework subject to lessons learned during a more detailed evaluation process.

We have to address both voice and data needs. However, we do not need to apply identical (to some level) schemes to voice and data. This certainly is the case with the AMPS (voice)/CDPD (packet data) combination. When we describe the Subscriber Demodulator Bank/ Modulator in more detail (in Section 3.6), we will start to build a case for having a common framing scheme for both voice and data. This begins a conceptual merging of an equivalent of AMPS and CDPD by retaining the key features of both systems.

Figure 9 is a top level system view analogous to what we have just discussed with respect to CDPD. The figure is the version of the Option #1 Architecture that we are discussing. There are no airborne ATM switches and thus absolutely no wireless ATM. All the switching on the airborne platforms is performed by routers. This has a strong advantage in that, at the current time, there are no wireless ATM protocols that could be used for the mobile WI. There have been a number of demonstrations of point-to-point ATM radio links, but these links are considered quite stable and are not similar to the point-to-point links that will be used to create the backbone. We recognize that one of the "features" of ATM is a better integration of voice and data; however, we feel that low bandwidth, wireless channels are not the best medium on which to apply ATM technology. (Option #2 examines a more wireless ATM-like solution both for the subscriber subsystem and the backbone.)

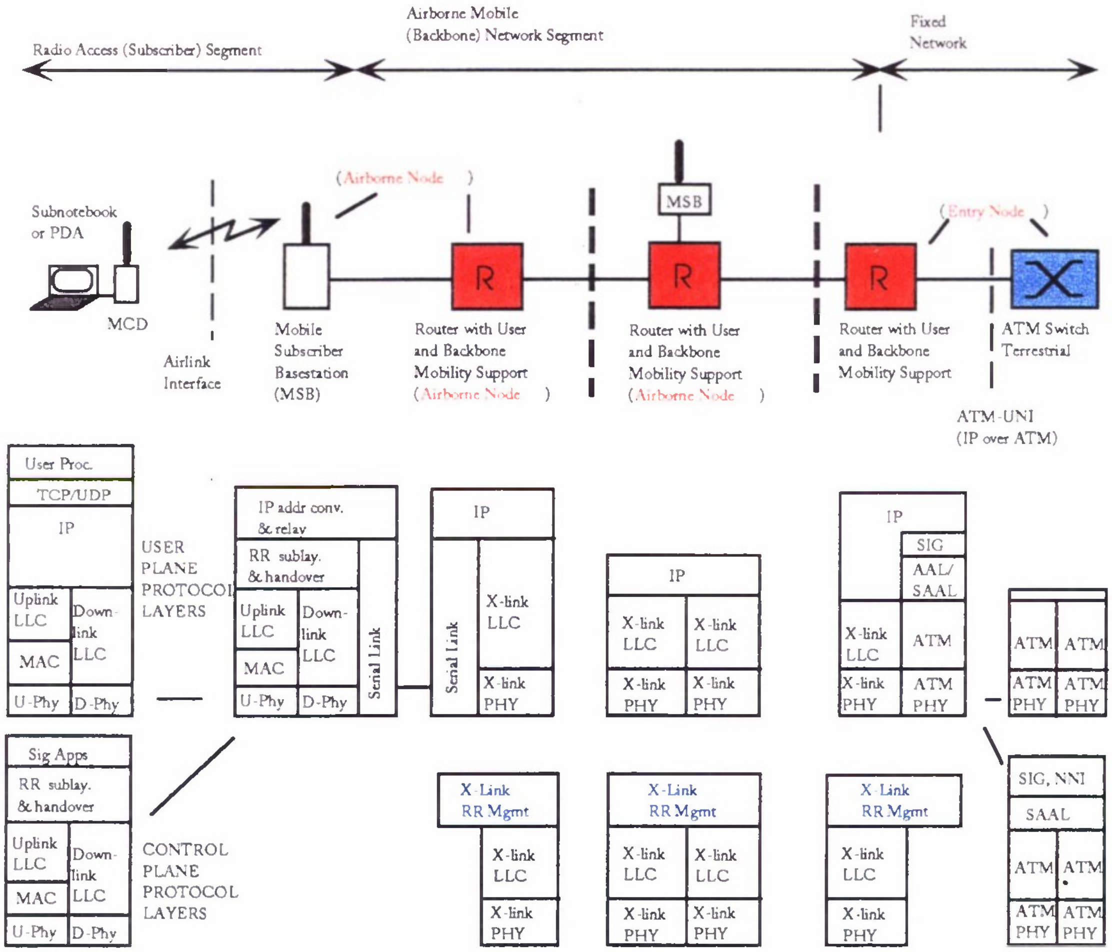


Figure 9. WI Option #1 with all IP router based network.

Since there are no on-board ATM switches, any traffic coming from other sources on the airborne platform must be first converted into IP traffic and then presented to the on-board router for relaying along the proper path. However, this may not be all that is required. We will come back to this point when we examine on-board routing especially with regard to mobile users.

The protocols shown here for the subscriber "airlink" differ from that in CDPD in a number of ways. We have not shown a separate SNDCP sublayer or a residual MAC downlink layer. These are minor differences and these will be addressed in time. However, we have taken pains to separate data and control / management layering more in keeping with telephony protocol standardization. However, the implementation really hides this separation. We have also combined a number of airborne subscriber subsystems (demod/mod, subscriber channel controller, signalling message handler, and subscriber routing function) into a single unit labeled Mobile Subscriber Base station. We will show the actual partitioning of the protocols among the subscriber subsystems when these individual subsystems are described in more detail.

It will become apparent that data and voice will be processed differently in the subscriber system design since their key QoS requirements are different. Data is intolerant of errors but will accommodate delay and delay variance, while voice will tolerate bit errors but is sensitive to delay variance. In the following two subsections we will add some more details on an efficient packet data system and then we will look at mechanisms of carrying voice traffic with modest changes that have little impact on the data system design.

[Variation on Architecture #1. Since the government is putting so much emphasis on wireless ATM possibilities, it is useful to provide one additional variation that shows wireless ATM in the backbone with the main data communications operation being IP over ATM. This is shown in Figure 10. Note that the subscriber system is still IP-based without the presence of ATM subscriber access schemes.

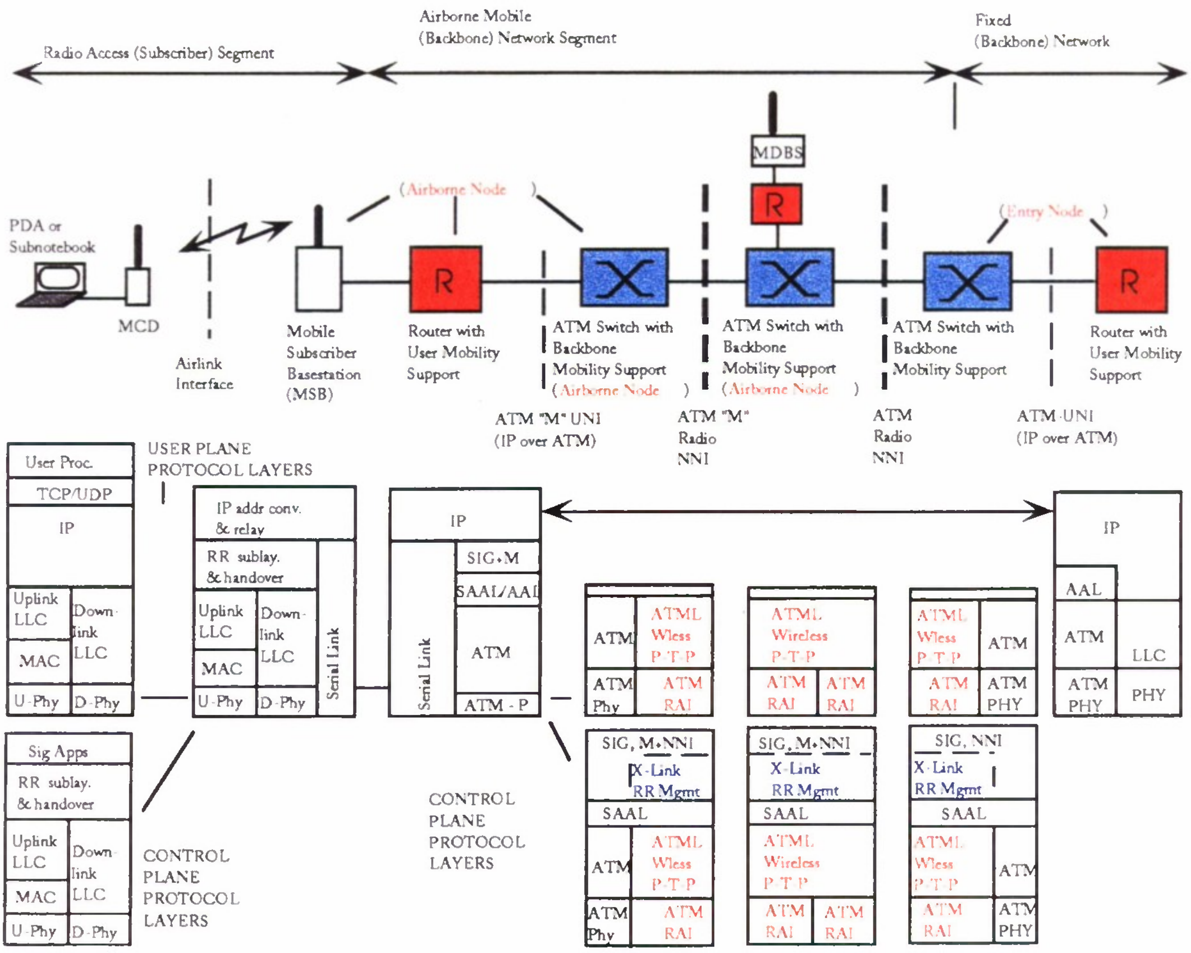


Figure 10. Architecture #1 variation with wireless ATM backbone.

It is useful to compare this figure with the preceding to compare the additional complexity. Not only do we need an additional switching subsystem on every airborne platform, but we also need special wireless ATM code that handles a movable backbone. This is a subsystem that the wireless ATM forum has yet to even consider.]

3.3.6.1 Data Issues

As implied by the overview and the description of CDPD, the uplink multiple access scheme will be a variation on CDPD's DSMA/CD. The goal was to avoid putting too large a scheduling burden on the base station controller. Moreover, there was a desire to eliminate the need for an uplink request channel to send up "fast reservation" requests and for a considerable amount of downlink capacity to communicate "grant" messages which include detailed uplink transmission schedules. Some downlink capacity is needed to provide the equivalent of channel busy tones and collision alerts. This, however, is a greatly reduced load.

Since this is similar to CDPD, there needs to be a microslot structure on both the uplink and downlink. The microslot interval is expected to exceed several msec (the exact value is subject to analysis). The duration is related to the maximum two-way propagation delay and some key processing delays which have yet to be specified. The data aspects do not place requirements such as fixed transmission units or uplink and downlink framing beyond microslots. These requirements will arise when looking at TDMA and voice but can be ignored in examining data. Our approach will be to look at each layer and list the services provided as shown in the figure above and note the commonality with those provided for the CDPD airlink.

Physical Level

It is assumed that the uplink demodulators and downlink modulator are assigned to specific center frequencies. There is a need to apply a spreading technique to the downlink and a despreading technique to the uplink. The key management system must supply the TRANSEC keys to provide both functions. (The MCD must have its center RF frequencies assigned to the proper modulator and the common demodulator and to perform the proper spreading and despreading functions again with supplied keys.) Downlink power control is managed in this layer.

The physical level also includes the modulation scheme and adds and processes structures needed for frequency, time, and bit synchronization. As in CDPD, one may also add packet burst trailers for the MCD transmitter rampdowns.

We will assume that the demodulator will process every uplinked "message" starting with the synchronization header and ending with its trailer. This is passed to the MAC layer for further processing. The recognition of the synchronization header indicates that the corresponding microslot is busy.

MAC (Media Access Layer)

This section is concerned with the MAC layer for a demodulator selected to support data only. Voice-oriented demodulators are treated separately. As in CDPD, we will assume a microslotted uplink and downlink subscriber framing. The microslots are on the order of a few msec. We will use 5 msec for initial purposes corresponding to roughly 500 bit times—assuming 100 kbps on uplink. We will revise this number in subsequent analyses but it should be approximately in the right range. Given that an uncompressed IP header is 320 bits–40 bytes, a 500-bit packet would be a relatively short packet. Microslots are the boundaries on which uplink transmissions are started and from which the base station tries to identify an empty microslot, a legitimate packet, colliding packets, and packets suffering unrecoverable FEC. The goal is to identify all three of the latter early so as to terminate their transmission and set up a retransmission.

In some contention-based schemes, once a single user has started its transmission successfully, it holds the channel until its transmission queue is depleted. This behavior is both good in that it fills the channel efficiently and bad in that this can result in an unfair utilization of the channel. In our scheme, this behavior must be inhibited. This can be enforced as part of the capabilities assignment process specified by the SCC. This system could specify a set of parameters based on user priority and message precedence. The parameters would be the maximum number of packets to be transmitted before yielding channel, the occupancy probability in p-persistent schemes, and the backoff retry intervals in both p-persistent and nonpersistent access schemes.

The uplink media access layer is more complex than that of the downlink principally because the uplink is multiple access (many to one) while the downlink is multipoint (one to many). Multiple access is basically a complex issue. Additionally, the MAC layer is where we choose to do forward error correction, which is required due to the high error rates of the wireless channel.

We will assume that when users transmit at the same time, the collision is detected by the presence of decoding errors. Decoding errors can also result from propagation degradation even when there is no collision. It would be desirable to distinguish between these two situations since it may influence the retransmission strategy. (This needs to be analyzed.) In any case, retransmissions will be at the MAC layer.

The downlink is handled differently. Collisions are not an issue since the base station controls the entire scheduling of downlink messages. However, propagation degradation can and will cause errors. In this case, error recovery should be made at the logical link level and not the MAC level.

3.3.6.2 Handling Voice Traffic

We have made the assertion that voice will be handled by segmenting the digital voice stream out of the voice encoders into transmission units with headers and routing these units over the subscriber subsystem in an analogous manner to data packets but with a QoS that is appropriate for voice. At this point, we want to discuss how one can transport voice over the subscriber system and leave the transport over the backbone system to a separate discussion. In our strawman system, although we are packetizing voice data for transmission over the entire WI, within the subscriber system we will still retain many circuit-oriented assignment characteristics. Circuit-oriented signalling is used to set up uplink and downlink time slots within the subscriber system only. If a voice packet passes onto the backbone, it is handled as a packetized voice packet and not as a circuit. It is important to keep this distinction in mind. However, the IP address in the packet is used for routing in the entire WI.

We have noted that one or more demodulators can be dedicated to supporting voice traffic. Since we are still interested in maximizing the number of simultaneous voice sessions, it makes sense to restrict the voice operations to a few kbps digital codecs. Modern digital codecs are being designed for wireless environments with data rates near 2.4 kbps (and after coding nearly 5 kbps). These low rates still produce voice quality with a MOS (mean opinion score) close to some of the older higher rate terrestrial cellular systems like GSM (at 13 kbps).

For voice operation, we will assume a TDMA scheme with uplink and downlink slots accommodating roughly 5 kbps voice streams. With suitable guard bands and signalling slots and with a maximum uplink burst rate, we could support greater than 16 TDMA slots / demodulator. [If all 16 channels are not in use, the spare capacity slots could be used to support data use. However, the data user must be careful in forbidding his uplink data transmissions from crossing TDMA slot boundaries. This is a different operation from data transmissions in pure data-oriented demodulators.]

One of the features of packet-oriented data networks is the wide range in packet lengths. Size differences can and do differ by over an order of magnitude. Long packets are, in one sense, more efficient because there is less overhead associated with the data enclosed with the packet. When there is no QoS delay or delay variability constraint, then mixing long and short packets does not matter and channel efficiency of use only improves with increased packet length. This is essentially the state of the global internet today with respect to data.

However, when average delay and delay variability becomes important, as it is in voice systems, then short packets become more attractive in one sense and less attractive in another. In many networks much of the delay and virtually all the delay variability comes not from propagation delay but rather from a processing delay. Some of the processing delay arises because it takes a fixed amount of time to accumulate enough voice samples to fill a packet. Generally, one wants to collect enough voice samples such that the packet header overhead becomes a smaller percentage of the entire packet. This is not a simple issue. Let us assume that the voice encoder produces an output rate of 2.4 kbps (which when encoded ends up near 4.8 kbps). In 10 msec, one has only assembled 24 bits of voice data (almost 6 bytes of encoded voice). In 100 msec, one would have 60 bytes of data. IPv6 packet headers are 40 bytes before we start adding option headers. Clearly, it is not efficient to send this header with each 10 msec assembly. The overhead percentage drops down significantly with the 100 msec assembly time so this is preferred. Figure 11 is useful to see some of the key issues on packetizing low rate voice.

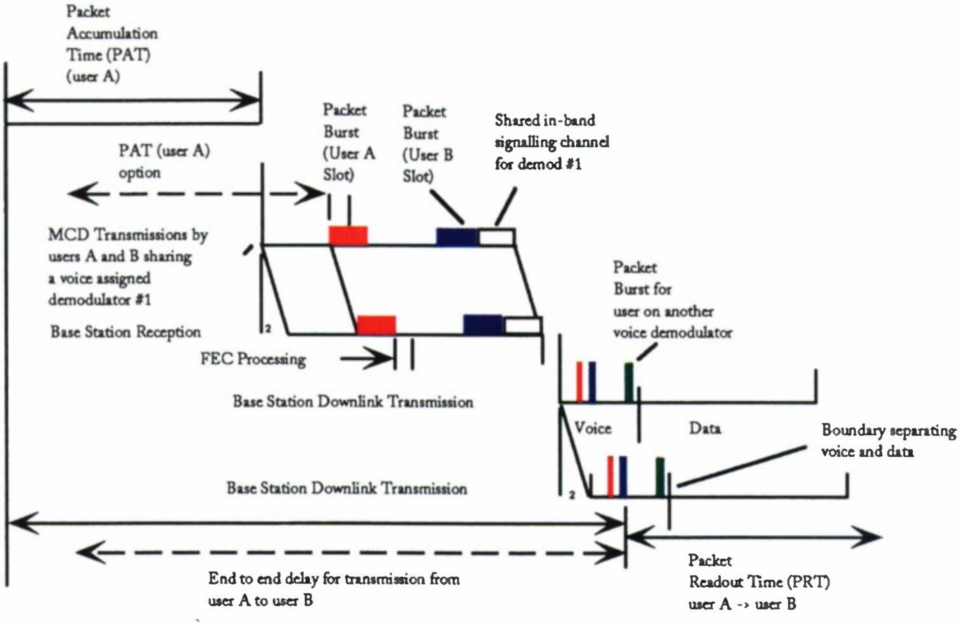

Packet Accumulation Time (PAT) (user A)
PAT (user A) option
Packet Burst (User A Slot)
Packet Burst (User B Slot)
Shared in-band signalling channel for demod #1
MCD Transmissions by users A and B sharing a voice assigned demodulator #1
Base Station Reception
2
FEC Processing
Base Station Downlink Transmission
Packet Burst for user on another voice demodulator
Voice
Data
Boundary separating voice and data
Base Station Downlink Transmission
2
End to end delay for transmission from user A to user B
Packet Readout Time (PRT) user A -> user B

Figure 11. End-to-end delivery times with low rate voice.

In this figure, assume that user A is transmitting voice packets to user B with both users being in the same beam. Assume that 100 msec of voice data are collected before being transmitted. User A and user B are assigned uplink and downlink TDMA slots as indicated. We should note that the downlink is organized so that the voice is transmitted before the data section. This is to minimize the overall delay and to reduce delay variability. The voice-oriented uplink demodulator frame is divided into equal time slots, each of which accommodates a single voice user. One or more slots are taken for signalling. {We should also note that the downlink segment called voice also has interspersed within it a number of words transmitted every microslot interval detailing the occupancy/collision history of the various demodulator channels. This history is used for the collision-based data access scheme. We will examine a more detailed and integrated framing structure later.}

The 100 msec forms a basic uplink and downlink frame structure. From this figure it is apparent that the end-to-end voice delay can be approximately twice the frame duration. If the voice accumulation times start exactly 100 msec prior to the scheduled transmissions, then the end-to-end delay can be somewhat reduced by about 50 msec on the average. This is shown as the dotted arrow.

Even if wireless ATM were available for the subscriber access, which it isn't, then the situation would not be significantly different. It would still require a cell assembly time close to 100 msec and since the CBR services generate empty cells even when there is no voice data, the rate at which one would generate cells is the same as one would generate packets as shown above. Thus the end-to-end subscriber delays within the subscriber subsystem would be essentially the same. The only difference would be fewer bytes of overhead attached to the cell as compared to an IP packet header.

Much of the accumulated processing delay can result from the time that a packet is kept on the set of output queues awaiting a scheduled transmission time. In the typical network node, there usually is a set of parallel incoming packet streams, each stream serving multiple users. In this node, the input stream of packets are processed and then routed to the appropriate output ports where they are put on the transmission queue subject to some scheduling policy. The difficult problem is to schedule the transmission of these packets on a set of outgoing channels so that a measure of QoS can be enforced. When a voice packet (or cell) is transiting the backbone, we want to minimize the time spent on the transmission scheduling queues. Both packets and cells would be generated every 100 msec and the added delay and delay variability depends on how one empties the queues either in a QoS-compatible router or wireless ATM switch.

There are two main problems associated with the statistical multiplexing. Scheduling is an inherently difficult problem and generally results in an NP-complete (non-polynomial) problem. Rescheduling is needed even for isochronous sources because of the appearance of new higher priority sources. It is well known that it is much easier to implement a priority-based scheduler when the "duration" of execution is a fixed quanta. Most statistical multiplexers work in this mode as well. It should be noted that all statistical multiplexers will produce delay variance associated with any particular user's data stream. However, the delay variance will be smaller for smaller scheduling quanta.

We should point our that if voice were the dominant information source in the WI, we would have used an end-to-end circuit-switched architecture. The architecture would be similar to many of the mobile terrestrial and satellite-based voice-oriented cellular systems. The advantage of circuit-oriented voice is that the scheduling problem associated with routers or ATM switches is eliminated. Moreover, total end-to-end circuit-oriented schemes can largely eliminate delay variance. Additionally, there would be no overhead due to packet headers since the routing was established on call set-up.

3.3.7 Quality of Service (QoS) Considerations

In most telecommunication systems, QoS is determined by the ability of the system to allocate the requested channel bandwidth for a long but unspecified duration. The probability of call rejection is part of the QoS consideration as is the reliability of the communication channels and switches. The channel is either operating or has an associated failed component that needs to be fixed. In the more advanced communications networks, QoS is contingent on the ability to either reserve "bandwidth" in the switching node or having a high probability of getting the desired bandwidth in a statistical sense by closely managing aggregate input traffic within the set of switching nodes. Generally, the paths connecting the switches, once allocated, deliver as needed.

The situation changes radically as the communication paths become liable to unpredictable behavior. One cannot guarantee a specific level of QoS even if some users have higher privilege levels. Let us assume that the communications paths are subject to multipath fading or unpredictable blockage as the endpoints of the channel move. In a typical cellular system, all the user paths from each user to the base station are different. As an extreme example, assume that a large group of low priority users have a benign path to the base station and the one high priority user has a path that is blocked or partially blocked by some obstacle. The question is whether it makes sense to assign additional assets to the high priority user to raise his QoS to what was requested. In this case, there is little one could do to help the high priority users other than to wait until the user moves out of the blocking zone. Clearly, reassigning resources assigned to the low priority users does not help.

When one has unpredictable links, a mix of isochronous and non-isochronous and a user group with variable priorities, the guarantee of a fixed QoS becomes very difficult. We will assume that the unpredictable links are mainly the subscriber links, and we will limit our compensation strategy to these end links only. This is to somewhat simplify the problem.

To illustrate this point, let us start with a single data session and assume that this session is operating just below its maximum rate for the promised QoS. If the subscriber channel becomes degraded, one could compensate by a combination of lowering the effective rate, interleaving, and additional coding protection. This would be a rate adaptation strategy and while providing data at the specified BER (bit error rate) would result in lower throughput so that the overall QoS (here throughput) is lowered. If, on the other hand, the session were operating below its maximum rate, then one could provide a more effective coding strategy without giving up the data rate but the margin will disappear, which is fine. Note that the user had to either rate adapt and provide more overhead in coding or to merely use up the excess margin by adding coding. In both cases the result is that we have the same average BER.

In this architecture, we are talking about shared subscriber channels. The ideal situation is to be presented with a user group whose individual dynamic traffic needs can be characterized, whose priorities are known, and whose QoS are realistically specified. Moreover, the link can be characterized sufficiently so that an allocated fading margin (or equivalent) will be supported. The subscriber channel controller is responsible for adding users to each shared channel up to some point where the QoS is anticipated to suffer. It is assumed that this allocation process has taken into account the expected loading efficiency of the shared channel under multiuser operation.

We have yet specified how one characterizes a user load. For the uplink, a qualitative measure would be the percentage of time that a user has captured the shared channel. Each user transmits in bursts and the normal assumption is that each user operates at the same burst rate. This high burst rate, of course, is far different from the user's average communication rate which scales down as the average duty cycle.

Now let us assume that the channel experiences some level of unexpected degradation over and above the allocated link margin for fading. How will this affect the user communications? Some aspect of QoS must suffer. The question is whether there are any simple ways of recovering some of the loss in QoS. Note that when the planned degradation levels are constantly exceeded, a “retry” transmission will probably fail as well and the result is that the user throughput could approach zero.

There are a number of methods to compensate for this added degradation, but they add complications to the design.

1. Let us ignore the fact that the users may have different priorities and let us assume that they are all data users. An acceptable strategy may for all the users to rate adapt (more encoding) and perhaps more interleaving. Their average burst duration would not change. All the users would suffer a loss in responsiveness but the link’s BER can be made closer to that of the standard budget allocation. One may not have the same ability to rate adapt for voice, and in this case, one would accept short voice outages.

2. Let us now assume that users have different priority levels and one can use this priority information to “bump” lesser priority sessions. This “bumping” would be based on both current traffic loading measurements and the relative priorities of users allocated to the demodulator channel. Let us assume that some of the higher priority users will not tolerate a loss in throughput and hence need more of the channel time for their added FEC. One would terminate lower priority users or could request that the lower priority users back off further in terms of their “burst” duration. This will certainly lower the throughput of the lower priority users yet retain the QoS of the higher priority users.

The complexity is starting to become evident even looking at these few cases. In case 1, we have made a dynamic change to the decoding scheme at the demodulator and requests have been transmitted to the users sharing that demodulator to all switch to a different encoding scheme. In case 2, we have again asked for a change in the coding for all the users sharing the uplink, but we have also asked the lower priority users to reduce their average access rate (less channel time).

In the last few paragraphs we have addressed the situation where the channel is “worse” than the allocated fading margin. What happens when the uplink is better than the allocated margin predicts? Clearly, we have surplus capacity potential that we cannot use. This is not an unusual case. In all cellular systems, the design is oriented around providing a specified QoS for the most disadvantaged terminal, usually the terminal at the edge of beam coverage. All terminals are additionally assumed to operate identically.

We should recognize at this point, for those terminals, say, directly under the airborne platform where multipath may be minimal, the throughput will be the same as any other terminal although the margin may be better. Similarly, in this architecture providing a small set of user terminals with directive antennas and higher transmit power levels may not help much under normal link conditions. They will still have the same throughput since we have not provided the capability for the uplink demodulator to dynamically change processing (via different encoding levels) from user packet to user packet.

However, one could place all the users with higher gain antennas and higher transmit powers in the same demodulator and apply a less powerful but uniform coding technique and thus improve throughput for the system as a whole. Incidentally, allowing demodulators to be operated differently is one of the benefits associated with Architecture option #1.

3.4 Subscriber Channel Controller (SCC)

It is useful to give a brief listing of the functions associated with the subscriber channel controller. Like the signalling messaging processor, this subscriber channel controller could be embedded with the Subscriber Router Function as one or more tasks. However, we will initially consider it as a separate device. It maintains a number of databases. As with most efficient database designs, special attention will be taken to avoid duplication of field information in the different database files. It should be mentioned that some of the databases need to accessed by the Subscriber Routing Function, and hence there needs to be carefully controlled access privileges.

One database is oriented towards unicast communications and thus contains records for all the users that are currently registered on the system as a whole. The set of all registered users is the global mobile user database. One of the fields in the record specifies whether the user is attached to this airborne platform or another (the value 0 indicates that the user is attached to this platform; other values represent attachments to other platforms). The local user mobile database consists of those records with this specific field. In this scheme, a user is only actively attached to one platform at a time, and this fact divides the user database essentially into those local users attached to this airborne platform versus those attached to the other airborne platforms (this provides a local versus WI wide view). If the destination IP address is outside the WI, there will not be a corresponding user entry since only directly attached users are put into the databases. It should be evident that as the mobile users transition between beams or attachments to a different airborne platform, the user's classification as local will be modified for on the databases serving the "left" beam and on the "joined" beam. This information needs to be distributed to all the mobile platform SCCs so that there is accurate information for all backbone routes.

One possible exception might be the legacy networks such as SINCGARS which are connected via a data device which, in turn, is connected to the MCD. In this case, we have a database file with the SINCGARS subnet, the associated WI MCD IP address, and the specific platform ID. This information is distributed to every router in the WI so that any user data can be routed through the WI to the correct gateway terminal and hence to the destination SINCGARS subnet.

Another database is associated with multicast IP addresses. The multicast address is the unique field. Within the associated record is a list of all the IP addresses connected to this platform which are members of this multicast group. For each of the IP addresses, there will be a pointer to a record where the fields that denotes its activity status, its currently assigned group (perhaps a beam ID), demodulator, and, if not implied, its modulator.

The subscriber channel controller is responsible for organizing the loading of the uplink channels as each user logs on, becomes authenticated, and then requests to be connected for specific communication services. The real task of the subscriber channel controller is assigning each new user to a specific communications channel based on a knowledge of the current overall assignments (and perhaps some indication of the current average demodulator loading). The assignment algorithm depends on the "access rights" communicated to the subscriber channel controller by the Authentication Center and the type of traffic request made by the user.

Most importantly, the subscriber channel controller will not be involved in the contention resolution process which defines the actual loading of each uplink demodulator. This sort of process was outlined in the section on CDPD and the WI uplink access scheme. This makes the relatively slow assignment process much less demanding than a "fast reservation" scheduling alternative. The overall goal is to provide each user with an efficient sharable channel with access rights compatible with the priority structure.

The subscriber channel controller role on the downlink may be minor since the downlink capacity should be considerably larger and downlink management is handled by the subscriber router.

We should note that because data is the dominant form of information and it is intrinsically bursty with a low duty cycle, the assignment of capacity to a user is a very difficult issue. In the strawman scheme, we will initially assume that a user makes a request for an uplink channel (which is implicitly shared with other users). If possible, a rough traffic descriptor would accompany the request. The assignment is for a user to a specific demodulator perhaps for a specified time period after which the user must again request an assignment. The request is coupled with the user's assigned priority and privileges, and with this information, a channel assignment is made out of the pool of demodulators. Packets generated by the user are transmitted with a Spread Spectrum code that guarantees that they will be correctly processed by the assigned demodulator. Packets are essentially uplink receiver directed.

Since these are shared demodulators, there needs to be a database uniquely identified by the demodulator ID containing a list of the active assigned users. Each of these users has either a default traffic descriptor or a requested traffic descriptor. The "sum" of these descriptors would often indicate maximum loading beyond the capability of the link, but clearly the actual instantaneous loading will be less than the maximum most of the time.

The second function is the dynamic management of the real time data transmission requests to guarantee that the uplink channels are loaded in a reasonable fashion. Note that we have differentiated the request for a user to be assigned to a sharable uplink channel from a request to be allowed to transmit a data packet over this shared channel. There are many schemes for coordinating multiple access to this channel from simple Aloha schemes to some complex reservation schemes. We have not yet decided on a choice of a specific multiple access scheme and thus defer the design details. However, if it is a relatively complex reservation scheme, one may want to remove some of the computationally intensive functions and put them into the block labeled Demod. In other words, some of the dynamic channel controller functionality is moved into the demodulator blocks. The demodulator would then not only provide the necessary standard demodulation scheme, but would also enforce the reservation algorithm. This would mean that the demodulator would trap reservation requests and generate reservation grant messages which would have to be encapsulated somehow and sent on a reservation grant broadcast on the downlink.

In this design choice, each demodulator manages its own multiple access scheme for the channel. The challenge is to minimize the overhead for the access coordination and to make a request only when there is "sufficient" data in the user's uplink queue. Note that the reservation grant would indicate when the transmission should begin and also indicate the amount of data to be transferred. This is clearly a dynamic TDMA scheme without fixed boundaries.

Decentralizing the real time part reservation system may not be the best idea since it implies a certain amount of inflexibility in moving different types of traffic across different demodulators. A fall back position might be for the main subscriber channel controller to periodically download to the demodulator the parameters of actual reservation scheme to be used. Let us assume that voice reservations will be handled completely different from data reservations. Then if, say, voice traffic requests increase significantly, one could move data traffic off a demodulator and reassign this demodulator to handle voice. This process would require coordination between the subscriber channel controller and the affected demodulator channels.

Once a user has registered with a channel controller, this entry is retained even if the subscriber drops its active connection. This is so that incoming calls and pages can be sent out on the out-of-band (OoB) signalling channel. Messages for an inactive subscriber can only be sent over the OoB signalling channel.

How does one delete the subscriber from the last channel controller visited?

1. If the subscriber is switched to a new aircraft, the entry is changed from local to reachable from another aircraft. The entry in the subscriber channel controller's database is modified and the aircraft's backbone router is updated with new reachability information.

2. If the user registered with another network (not the WI), then that network's FA will inform the user's home network which will register a change in the WI FA. The WI's FA will then send a delete message to all the backbone nodes.

3. There may be a last resort delete policy. This is to account for the situation where the user remains inactive for extended periods. One wants to prune the lists in the subscriber channel's controller and in the backbone router. Perhaps send out a "where are you" message and if not recognized within a specified number of days, the entries are deleted and the home network becomes the default location.

4. When the subscriber has knowledge that it is moving away, there should be a formal message that this is going to happen. This is sufficient information to update the WI tables in all the routers.

3.5 Subscriber Router Function (SRF)

It is important to recognize that in the subscriber subsystem there are two units that handle routing—the SRF and the Backbone Router (BB-R). The SRF accounts for the mobility of the users whereas the BB-R has no concept of user mobility but it must accommodate itself to mobility of the backbone. The physical backbone connections are established and managed by the backbone connection manager. The routing over the backbone is determined by the BB-R so that user packets are delivered first to the proper airborne platform. The associated SRF then makes sure that the user packets are sent out over the proper downlink. The assumption is that there is only a single high capacity downlink-per-downlink beam which is shared by all users under that beam. Another assumption is that the airborne antenna's uplink and downlink beamwidth for each beam is equal so that we have equal coverage areas.

As we noted earlier, the (central) subscriber channel controller and the signalling message controller may be special tasks in a candidate implementation of the SRF. Aside from these possible functions it has its own unique responsibilities. End user routing is certainly a major function. Another is the demultiplexing and multiplexing of the subscriber traffic. The beam handover management process is another of its major functions. Separate sections will be devoted to each of its major responsibilities.

3.5.1 End Link Route Determination

Let us examine the various routing cases with respect to subscriber traffic. For the simplest case, consider the case where both users are covered by the same beam. In this case, each user's packets are sent to the SRF which uses the local database to match the destination IP address to an active user for the same beam. Assuming that the destination address is local, then the router will queue the message for transmission over the downlink. The backbone router does not participate in the information exchange. The two terrestrial users merely pick off those packets with destination IP addresses matching the receiving user's home IP address. (Actually, one picks off packets with a matching machine address as explained in the section on Demodulator Bank / Modulator Subsystem.)

If the two communicating users are under the same aircraft but in separate beams, then user uplinks in one beam are routed through the SRF to the TDM channel serving the other destination user. Again, the BB-R is not involved.

Now consider the case where two communicating entities are under different airborne platforms. In this case, an uplink packet will have a destination IP address which is not in this platform's local database. Let us assume that the destination user had registered so that his home IP / current airborne platform IP pair is in the global database. In this case, the SRF could add an extra IP header with the destination user's associated airborne platform IP address as part of the header. This packet is sent to that airborne IP address, and at that point the destination SRF would remove the second IP header and expose the destination IP home address. This remaining packet segment has its logical link level subheader added on and then sent on the downlink as expected.

Finally, let us assume that a Warfighter user wishes to communicate to a user totally outside the mobile WI system. In this case, the user sends an uplink packet with the destination user's home IP. This home IP certainly does not match either the airborne platform's local database or its global database containing all the registered users connected to any airborne WI platform. The strategy is then to send this packet to the BB-R, and since the IP address corresponds to a location external to the mobile WI, it should be sent to the gateway router at the mobile support node for external distribution. In the implementation, the routing can be aided by having the SRF encapsulating the packet IP address of the gateway which then is used for routing within the backbone. At the gateway, this IP address subheader is removed leaving the destination IP header, which then is able to find its way across the external network to the final destination.

At this point, we need to review how we intend to handle address mobility within the mobile WI subnet. We postulated a foreign agent function in the specially augmented router located at the mobility support node. We implied that this foreign agent "knew" how to route a packet destined for a mobile user connected to this mobile WI. Remember, this mobile WI is an autonomous system and its routing can be selected in a unique way appropriate to the peculiarities of this subnetwork. Let us now be more specific.

Each backbone router has a set of I/Os attached to bidirectional channels that connect to other backbone routers and also a connection to its local subscriber router function. Each of these I/Os has an associated fixed IP address within the mobile WI network address space. The backbone router is assumed to build up dynamically a database of path segments to all the other active backbone routers on the mobile WI subnet subject to the information supplied by the Backbone Connection Management unit. This unit will provide a list of only active backbone router IP addresses, not mobile user IP addresses.

3.5.2 Demultiplexing Function

A traditional router normally has one I/O port associated with each connection to a different network or to another router. These I/O ports also have their own unique associated IP addresses. The SRF is not a traditional router but behaves more like a host connected to a traditional router; in this case the traditional router being the backbone router. We should point out that the SRF ports provide links to the demodulator outputs and to the modulator inputs. None of these ports connecting to the demodulators or modulators have IP addresses. The associated IP addresses are those located at the end of the links: the MCDs. So these ports are more analogous to a host's terminal ports. Of course, the association of a particular user with a "terminal port" is really an association with the end user's IP address.

The SRF does have a single IP address and that address is associated with the link to the backbone router. We will not initially consider the Signalling Message Handler as having a separate IP address but will assume that any messaging destined for it will be addressed with the SRF's IP address.

3.5.3 Beam Handover Processing

Beam handover is a complex process involving the MCD, the airborne demodulator, the SRF, the SCC databases, and ultimately the routing tables in the Backbone Routers. The details of the beam handover have yet to be worked out.

3.5.4 User Traffic Scheduling Support

The SRF is where the local subscriber's uplinks are connected to the shared downlink. Thus, in the suggested contention-based system, as each demodulator reports in on occupancy and collisions in each uplink microslot, the SRF must create an aggregate channel report (one or more bits) to be transmitted down as one or more 32-bit words in a corresponding downlink microslot. We will discuss uplink and downlink framing issues later with respect to the Demodulator Bank / Modulator discussion.

3.6 Subscriber Demodulator Bank / Modulator (SDB/M)

In describing the Demodulator Bank / Modulator system, we will go beyond the issues of selection of an appropriate waveform and spread spectrum choice, synchronization design, modulation and demodulation techniques, and coding strategies. We will try to look at the functions that will be assigned to this system. These include strawman designs to provide a sufficiently efficient solution for uplink access balanced against downlink capabilities. We will look at the mechanisms that enforce the access policy sent from the Subscriber Channel Controller which in turn was obtained from the Authentication Center that was "set up" by the tactical planners.

The design of the SDB/M must go hand in hand with the design of the MCD. These will be drastically different designs, but there is a close synergistic design relationship between the two. There is tremendous differences in the responsibilities of each. To a high degree there is a strong master-slave relationship between the two. An MCD is designed to operate (in a sharing mode) with connection to one demodulator at a time and share the single downlink with all the other users under the same antenna footprint. The SDB/M with its 32 demodulators and a single modulator is assigned the role of coordinating the communications function of deconflicting the communications.

The choice of a multiple access uplink algorithm and shared downlink multiplexing scheme have yet to be made. Therefore, one is restricted to addressing key issues and reviewing some implementation options which are only meant to clarify some of the issues but certainly are not implementation-specific directives.

In the proposed access scheme, the implication is that a user, after a request for service, is assigned to a particular uplink demodulator by the Subscriber Channel Controller for the duration of that transaction session. Generally, the user will be sharing this demodulator with other users. One of the strengths of the assignment of each user to an individual demodulator is that one can provide different types of service by a user-to-demodulator assignment. For example, the

Subscriber Channel Controller could assign a subset of demodulators to support voice. The uplink channel can be broken into N TDM slots with each slot being assigned for an active call for the duration of the call. [One or more of the TDM slots may be used on a contention basis to pass uplink signalling for all the users sharing this demodulator.]

A demodulator assigned to support data traffic could be operated in a completely different manner. Consider bursty data communications as being a highly used operational mode. Assume that these users generate bursty traffic in an uncoordinated manner. The challenge then is to provide an uplink channel sharing scheme that is reasonably efficient given that one cannot predict when a user generates information to be transmitted.

Similarly, data is arriving at each SRF from multiple sources for the purpose of being aggregated and sent down on a shared downlink modulator. If quality of service or priority were not a consideration, one would merely create a single queue of "information units" and concurrently empty the queue via transmission as fast as the downlink capacity will allow subject to link quality control procedures. We have just used the expression "information units" since we do not initially want to restrict ourselves to making the association "information unit" equals a "packet" We will look at some of the advantages of having "information units" different from a "packet" in a following section on segmentation.

The following sections will address different aspects of the efficient utilization of both uplinks and downlinks subject to imposed QoS conditioned by traffic types and user priorities. It should be noted that the uplink channel assignments by the subscriber channel control contributed partially to setting a viable QoS. The remaining control of QoS on the uplink is determined by the quality of the propagation channel and the uplink multiple access scheme.

One assumption that has been made is that the MCD has generated IP packets, and even if these packets are temporarily segmented for uplink access, these segments will be reassembled in the uplink demodulator before being passed to the SRF. Thus, the SDB/M will support the protocol layers up to a subnet layer below IP.

3.6.1 Physical Level Considerations

3.6.1.1 Spread Spectrum and Frequency Management

In the WI, we are interested in frequency reuse and the possibilities of AJ (anti-jam) support. This suggests utilizing some type of spread spectrum (SS) technology. At this time we will defer the particular choice of SS, although our initial preference will be some form of frequency hopping. At this point it is sufficient only to assume that the Demodulator Bank / Modulator is provided with a set of SS keys, each of which is used to generate a pattern which is to be combined with data (actually both signalling and user data).

In this strawman implementation, the group of "j" users which are assigned to an uplink demodulator all share the same SS (private) key as does the corresponding demodulator. These j users then would directly interfere with each other whenever there is overlap in their uplink transmission times. A different group of "n" users would be assigned to another demodulator and given a different SS key. This group can self interfere with each other by having overlapping transmissions; however, these transmissions will interfere with the other group in a less severe manner. Even if users assigned to one demodulator transmit simultaneously with users assigned to another demodulator, the cross demodulator interference is reduced to essentially much lower white noise. The larger the SS bandwidth, the lower the resulting interference.

For the downlink, all the users under the same beam are assumed to share the same downlink SS key. There is, of course, no absolute reason why they need to share the same key, but multiple downlink keys are more difficult to manage. (The exception would be the out-of-band downlink logical channel, which could be an allocated TDM slot.) We recommend a different key to limit the illegal user from performing traffic analysis on the downlink. The assumption is that it will be easier to obtain illegal access to an out-of-band signalling channel compared to obtaining access to a data channel or in-band signalling channel.

For discussion purposes, let us assume that the WI is able to obtain 10 MHz of bandwidth for both the uplink and downlink subscriber allocation. The uplink data burst rate is assumed to be approximately 100 kbps and the downlink about 1 Mbps. With a bandwidth utilization of about 1 bps/Hz, one would have a downlink processing gain of about 10 dB and an uplink processing gain of 20 dB. Uplink jamming is clearly a more serious threat, and one can add to the jamming margin by operating at lower uplink data rates. One can obtain additional jamming margin via the use of a multibeam antenna that can place antenna nulls on jammer locations. Jamming protection will be a separate analysis. It should be mentioned that for data, the natural asymmetry in client-server data rates will somewhat offset a more serious jamming attack in that the server will produce most of the information transfer and many servers will access the airborne platform through a narrow beam high gain antenna. [A more recent link budget suggests that the supportable uplink burst rate is closer to 64 kbps and the downlink closer to 1.544 Mbps.]

A following section will indicate how to use multiple downlink keys to help realize the logical out-of-band modem signalling functionality without adding a different class of hardware.)

3.6.2 MAC Level Considerations

In Section 3.3.6, we started to introduce an implementation that adopted some of the concepts behind CDPD and yet accommodated voice needs. To produce an integrated scheme, it seems wise to propose a subscriber subsystem uplink and downlink framing structure that can accommodate both data and low rate voice needs adequately. While the end-to-end attributes are IP packet-based, within the subscriber subnet, one can take variable length packets and reduce them to a fixed segment size for reasons of efficiency and more predictable quality of service, and then reassemble these segments into IP packets for transmission to the backbone.

3.6.2.1 Uplink and Downlink Framing Considerations

At present this section does not have any fixed values until we perform a detailed analysis. However, we can indicate some approximate values. We have postulated voice operations at 2.4

kbps with added encoding and overhead bringing the rate close to 5 kbps. Framing is determined principally by voice considerations. One chooses the frame by balancing the delay invoked in accumulating sufficient voice samples to be packed into a transmission unit against the efficient unit size for transmission. Voice frames should probably be at or under 100 msec on both the uplink and downlink. Within an uplink channel supporting voice only, we would then divide this 100 msec frame into less than 20 TDMA slots (assuming an uplink burst rate of 100 kbps). Some TDMA slots are used for non-voice purposes.

In the section on data access schemes, we discussed the idea of a microslot which is on the order of a few to several msecs. This slotting is defined for the start of data transmissions on the uplink and as a downlink report slot indicating uplink microslot occupancy and collision status on a microslot-by-microslot basis.

Overall timing implementation would be simplified if the voice frame duration is made equal to the microslot time. We also showed on the downlink a movable boundary between the voice and data domains. This was an attempt to reduce delay variability on voice.

3.6.2.2 Uplink Access Scheme

At this point, it is sufficient to state that we will be investigating a set of contention-based multiple access schemes when the airborne base station reports back collisions. The entire scheme has to allow for servicing users with different user priorities.

3.6.2.3 Downlink Access

The downlink is a considerably simpler problem. However, the downlink scheduler must still guarantee that the isochronous services are provided adequate delay and delay variance bounds without severely compromising the data throughput. Additionally, the downlink scheduler has to perform this task and at the same time provide priority-based services. Both of these are challenges that the commercial wireless services have yet to address.

More details on the uplink and downlink access will surface in the on-going strawman implementation design studies.

3.6.3 Realizing an Out-of-Band Modem

There is one situation where, say, two downlink keys can be used in a productive manner. Let us assume that one key is part of the hardware / smart card of every MCD. When activated with a legal user password, the MCD using the shared key is able to read part of the shared downlink stream. This part (for example, a dedicated time slot in the downlink frame) provides both downlink synchronization capability and a downlink out-of-band signalling channel. The rest of the downlink stream can be read only if the MCD is provided another key following a completed user authentication process.

The provisioning of a downlink time slot for out-of-band signalling and synchronization fits well with the desire for the MCD to actively listen for pages and call alerts on a low duty cycle basis to minimize MCD power needs.

Presumably another shared uplink key in the MCD could provide access to a set-aside demodulator which aids in the uplink acquisition process and provides the uplink out-of-band signalling channel.

This scheme provides the logical out-of-band signalling modem functionality that we have referred to earlier. We did not have to introduce any hardware which is unique and this simplifies the design. We will expand on this when discussing synchronization techniques.

3.7 Backbone Router (BB-R)

Architecture option #1 was designed with the intent of allowing the backbone router to be essentially a standard router with a few exceptions. The following addresses some of the exceptions.

If there is a main ATM switch on each airborne platform, we assume that the operational mode is going to be IP over ATM. The assumption is that the users generate and accept IP packets which are passed between the SRF and the associated BB Router. This BB Router then has another attachment to the ATM switch and all traffic between the different airborne platforms is ATM cell based.

It there is no ATM switch aboard, then the router could have as many I/Os as there are external X-link antennas plus additional I/Os to the SRF and the SCC. The details of the MCA design will specify how many cross-link ports are needed if there is multiplexing (routing) performed in the MCA. This remains to be specified.

3.8 Backbone Connection Manager (BBCM)

This topic will be covered as part of option #1 MCA functionality for "neighbor discovery," determination of an active backbone topology, and integration of this active backbone either into an ATM switching fabric or into the router when ATM is absent.

3.9 Airborne Subscriber System Design Environments

Many of the subsystems in the airborne node (and also at the ground entry node) can be implemented entirely on high performance PCs. Clearly, some of these PCs may need special I/O cards yet to be specified. However, the overall hardware design task for these subsystems are going to be minimal.

The major design tasks will be centered on developing protocol software, much of it is oriented around telephony and networking protocols. However, these protocols largely are not standard protocols since a number of IPv6 protocol components are only now coming out for evaluation, and the standard TCP/IP stacks are useful only as points of departure for design considerations. The telephony protocols have only rough similarities to the types of protocols developed for GSM and other cellular-based telephony systems, and these must be developed from scratch. However, there will be many similar efforts that one can "borrow" code or designs from, and the key point is to use a development platform that is amenable to hosting experimental networking code and to run performance simulations on.

It might first appear that the Windows NT might be a good OS to start with. It is an excellent system on which to host many applications; however, because this is a closed system, it is the wrong system to use if one is writing low level drivers and is desirous of changing some of the low level networking code.

A preferable approach would be to start with an open OS such as Linux, particularly with the Mach-based version, and to develop code on this open platform. This OS is quite stable and is becoming the choice of many researchers in the networking world. The considerable advantage is that all the source code is available. Moreover, there is a wealth of networking code (some good and some bad) which is easily available. Finally, any code developed could be left in the public domain or as government property. This last aspect gives an open OS a tremendous advantage since its influence extends from the networking considerations through the distributed applications environment.

4. Subscriber Terminal (Mobile Communications Device—MCD)

4.1 Overview

In the section on the Demodulator Bank/Modulator System, it was mentioned that its design was intimately linked with that of the MCD. It is a parallel and closely coupled design which also has to address a number of common problems such as waveforms; an affordable, power efficient spread spectrum technique; a channel sharing implementation; switchable uplink sharing modes (voice versus data), etc.

The MCD here is considered as a single mode device; operating as a subscriber terminal with mainly airborne base stations. There may be a reason to incorporate a ground operations mode (as an SUO-compatible packet radio or as a radio that accesses a mobile terrestrial base station which could well use WI access techniques).

The MCD is considered a uniform device. In the initial design, it is hard to see how one could easily exploit the additional MCD types (with higher power levels or directive antennas). Such added capabilities at the user level would require commensurate changes in the Demodulator Bank/Modulator System to fully exploit this added potential.

The MCD provides the ability to switch between voice and data operations with only a few manual steps. One can even concurrently operate voice and data if this feature is designed into the DB/M system.

4.2 Design Constraints

As stated in the main body of this report, power management is a prime consideration in the MCD design. The design will be based on those technologies developed for cellular phones rather than those developed in association with "software" radios. The main power source has to be batteries, although this does not eliminate running off a vehicle's power systems.

There is a great deal of modem and protocol complexity and much of this cannot be executed in DSP chips because of the power budget. While prototypes can use DSPs and general RISC chips, the operational versions will use ASICs and a very low powered RISC CPU (limited to the 100 mW region). Efficient RT kernels with a small memory footprint will be needed, and only a few data applications should be resident in PROM. One of these would be a Browser which can download applets which greatly expands the number of runable applications. The data handling aspects of the MCD is where this device differs from the normal cellular phones with modem-attached PCs.

4.3 MCD Network Protocols

The MCD is the most challenging subsystem with respect to protocol development. It is involved in many of the protocols involving signalling interactions and in virtually all the protocols dealing with user information transfer. There is considerable complexity associated with signalling given that the term here is extended to include power management, uplink and downlink synchronization, user registration, authentication and security protocol support, etc.

In Figure 12 we show the user terminal as consisting of a voice device (the radio part) and a data device (a notebook computer). However, we are intending to allow flexibility on the choice of data devices. In the main section, we mentioned that the data device could be a subnotebook computer, a notebook computer, or a personal data assistant. These will undoubtedly be commercial equipment and the only modifications for these data devices could be unique PCMCIA cards and software drivers associated with these cards. The radio would be unique and would contain most of the mobile WI-unique communications and networking code.

In Figure 12 we have partitioned the protocol stack across both devices containing the WI-unique code to the radio. Again we should note that we do not show UDP or the IP support protocols on this diagram, but they are implicitly included. Ideally, one would like to process incoming and outgoing voice calls without activating the notebook computer. This is a result of the notebook computer having a high battery drain in comparison to that of a radio. If the notebook is inactive, one would need to retain enough of the signalling control processing in the handset to receive a call alert. Similarly, one would want to run the voice protocols in the radio (voice instrument) that carry the voice traffic, again with the notebook off. Initiating a call presents a minor issue. In this system, the callee's number is not a telephone number but rather an IP alphanumeric string (dpwhite@wi.mil). This could be done through the notebook or PDA interface, but this would require the notebook to "boot-up" and could be time consuming in adding to the call set-up time, making it unacceptable. The option is to have a set of well-known numbers mapped into several digits that could be entered from a keypad on the radio (a "speed dial" feature). For full directory services, it would be necessary to activate the notebook.

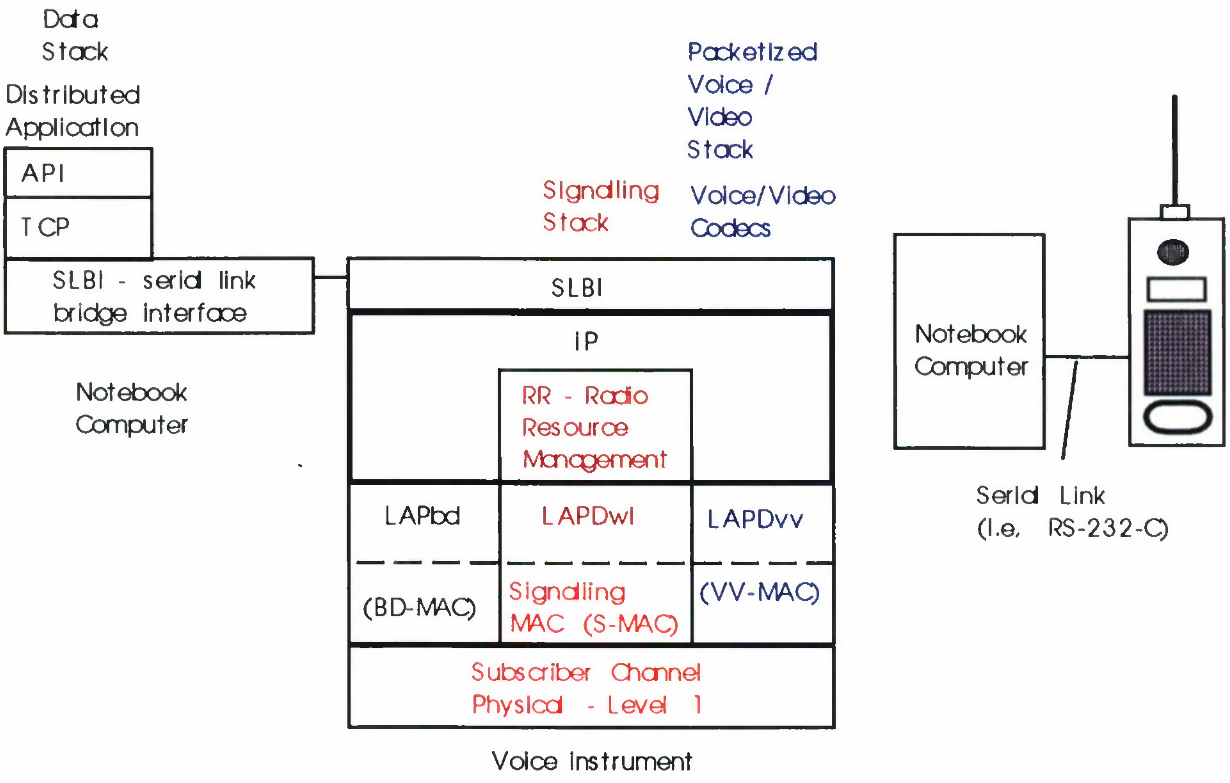


Figure 12. MCD protocol partitioning.

The handset can be put in a standby mode in which it periodically passively synchronizes, perhaps actively registers, and is ready for incoming calls or pages. It is normally operated in this standby mode, but strict power control may be enforced in that it may only "listen" at low power levels in specific time intervals. These are the intervals where call alerts and pages are sent. Listening means that it looks for packets with its unique machine address which we have called a MAC address due to its analogy with LAN terminology. Rejecting packets via the MAC address filter means that one need not process packets at higher levels that are not intended for this user. This is another power saving step.

Receiving a page may result in the page message being queued in the handset until the notebook computer is activated. There will have to be a page indicator on the voice instrument that is a prompt for the user to turn on the notebook device. After boot-up, the page (or short message) will be displayed; alternatively, it could be displayed on the MCD LCD display.

With the partitioning on information services across both the notebook computer and in the radio/voice instrument, it would seem logical to place the end user security device in the radio. One ordinarily thinks of a Fortezza token as being inserted in the notebook or PDA, but this would not handle the voice stream. One could use another encryption algorithm for voice and embed this in the radio, but this is starting to complicate security since one would like to handle all packets in the same way. It should be clear that one has to closely examine the functionality that needs to be installed in the radio. Initially, it appears that virtually all the communications and security functions need to be put in the radio as well as the voice codec. The notebook has a very limited communications role.

The Warfighter's MCD can also act as an interface to legacy IP-based terrestrial networks. One obvious candidate is the SINCGARS IP network which is an augmented SINCGARS function. In Figure 13 we show one possible protocol structure. (The voice stack is crossed out since it is not relevant to the SINCGARS data transfer task.) (SLBI is the serial line bus interface.) In this figure the notebook computer only has to execute the SINCGARS subnet adaptation layer code. This is subnet-specific code that fixes any mismatches that SINCGARS presents with respect to normal IP connectionless services. If there are none, then this would be a null sublayer and the SINCGARS radio could be connected directly to the radio terminal. We will leave the attachment to the notebook since this architecture would be easier to maintain for other legacy networks that may really require some special adaptation.

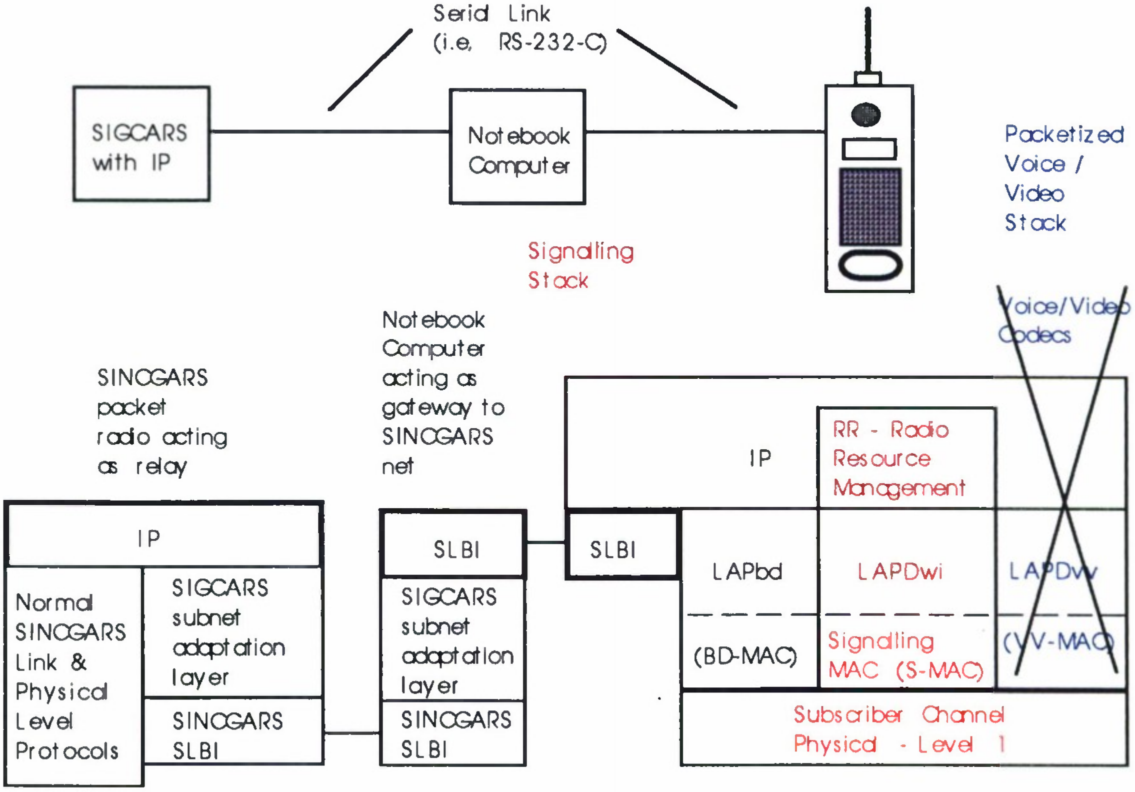

Serial Link
(i.e. RS-232-C)
SIGCARS with IP
Notebook Computer
Packetized Voice / Video Stack
Signalling Stack
Voice/Video Codecs
SINCGARS packet radio acting as relay
Notebook Computer acting as gateway to SINCGARS net
IP
Normal SINCGARS Link & Physical Level Protocols
SIGCARS subnet adaptation layer
SINCGARS SLBI
SLBI
SIGCARS subnet adaptation layer
SINCGARS SLBI
SLBI
IP
RR - Radio Resource Management
LAPbd
LAPDwi
LAPDvv
(BD-MAC)
Signalling MAC (S-MAC)
(VV-MAC)
Subscriber Channel Physical - Level 1

Figure 13. Interworking with legacy networks via the Warfighter's MCD.

4.4 Commercial Designs as a Starting Point

There is no question that the design of an MCD with the attributes listed in Section 3.5.2.1 is very difficult. There are many individuals and government organizations that believe that the commercial world is best suited to perform this design. This would be true if the commercial world clearly saw such a development as resulting in an almost immediate or near-term revenue stream. The terminal would only be part of the investment and it is only the most visible part of the entire system. Generally, the commercial world is not set up to develop technologies that require extended development without a clear picture on their return on investment. However, if the new system can be retrofitted into one of their own products or systems, or into a nationally or internationally approved working system, they would be much more inclined to make the investment.

There is also some misunderstanding on how advanced our US handset technologies really are. Most of the design experience is with analog for AMPS. However, as a result of the newer TDMA and CDMA terrestrial and satellite cellular systems, there is a considerable amount of investment in handsets for digital schemes. However, their design focus will be on providing voice services for future digital cellular and PCS systems where a guaranteed revenue stream is almost certain. Data will be accommodated over voice circuits, and although it may be packetized, it will not be an efficient use of resources. It is questionable whether industry can be convinced to invest heavily with or for the government on a packet data system handset specifically designed for efficient data operation with a mobile backbone. Moreover, since handset design is a carefully protected technology for competitive reasons, one cannot expect that the government will have much insight into the design details. Intellectual property will become a major issue.

It is important to realize that commercial cellular and PCS digital communications systems are split between FDMA/TDMA and CDMA realizations. The FDMA/TDMA systems have absolutely no interference protection other than international spectrum management. Providing intentional interference mitigation (AJ) would require a complete RF radio redesign, and this alone makes current designs highly vulnerable. CDMA has limited interference rejection since it is intrinsic to its design to make other (mutually interfering) subscribers look like white noise. However, it is the most promising approach subject to significant modifications to make it useful in tactical situations. (One still must address whether frequency hopping might be a better approach! CDMA with Direct Sequence spreading may not be the best solution for the WI since the WI needs to address military interference threats rather than normal interference.)

The US is clearly the leader in CDMA systems. The principal companies are looking to expand interest in their CDMA technologies in cellular terrestrial and satellite systems as well as the newer PCS systems. There certainly is an interest in packet-oriented data services, but this interest lags far behind their drive to provide effective voice services where the revenues are more certain. There is a standards group for a packet service for IS-95; the main drawback to putting a packet data service onto IS-95 is that the current constraint is in conforming to IS-95 channelization. The WI constraints are not the same since there has been neither a total spectral allocation nor detailed channelization. In any case, packet services over IS-95 are not an immediate priority to any company.

There are also many ISM band special developments taking place. We should carefully assess any promising technologies and see if any of it can be readily integrated into an appropriate mobile WI terminal, perhaps operating at another RF band.

4.5 Design Methodology

All of these modern cellular radios require a heavy investment in software/firmware development partly because their networking role is so complex and partly because many of the traditional radio functions are more economically executed in firmware. However, the military radios built until recently are heavily hardware-oriented with only a small firmware component. The primary reason was that A/D and D/A technology was not sufficiently advanced to take advantage of digital processing. Additionally, most of the radios were built to support voice and did not require the complex protocols needed for data (or even sophisticated signalling for voice operations). However, the A/D and D/A technology (as well as signal processing chips) has advanced in the past several years and the design space has enlarged.

Software radios like DARPA's Speakeasy showed the flexibility of up-front digitalization in that it allowed via optional firmware to emulate any number of military radios. As such, it can be positioned as a reprogrammable substitute for any of the current radios. However, Speakeasy as a testbed for a cellular-like Warfighter MCD has major limitations. First, its initial weight is in the area of 30 lb which is not a good start if it is to be shrunk to considerably less than 3 lb. Some of its flexibility in terms of I/O is a liability. Even more significant is that the programmability is via DSP (4 TI TMS320s). We are anticipating that the flexibility in the WI handset has to be both allocated to reprogrammable hardware and to DSPs with perhaps the reprogrammable hardware being dominant. If this is true, any already-developed DSP code will be largely irrelevant. Additionally, the considerable amount of unique code for the mobile WI has yet to be developed, and Speakeasy does not offer any development advantages compared to a newer development platform that is closer to the state-of-the-art.

The design of ultra-compact but complex communications subsystems like digital cellular phones is requiring a new design philosophy which increasingly relies on new powerful design methodologies and tools. The designs are neither hardware nor software oriented but are being considered as **hardware/software codesigns**. This is an important point. One recognizes that VLSI design and simulation is starting to require as many software skills as hardware skills. The same is true with FPGAs. In fact, the overall specification is starting to look more like a total software specification which can be mapped into either hardware or onto a software-oriented DSP. Codesign is a recognition of this fact and furthermore brings out the possibility that the partitioning being a hardware element map and a DSP map is not fixed but rather flexible and interwoven and only determinable by simulation.

Since we are concerned with weight and power, the evaluation via simulation of many of the hardware/software tradeoffs have to be done in detail before realistic weight, power, and performance assessment can be made. This significant codesign process must be fairly mature prior to committing to an implementation. Whereas once we had a loose collaboration between a systems engineer, a software designer, and a hardware designer, the codesign forces the designer to assume the roles of all three to some extent. These design capabilities are now appearing in industry, and should industry be convinced to devote resources to a compact MCD design, there is little doubt that it could be achieved.

5. Routing in Mobile WI

In the main text we mentioned the advantage of configuring the mobile WI as an Autonomous System which would allow it to use its own internal routing protocols. This has a number of benefits for mobility as we shall see in the next section for multicasting in a mobile environment. In this appendix, we are only providing an outline of how the mobile routing function can be accomplished. It should also be recognized that the size of this autonomous system will be relatively small; perhaps no more than a dozen backbone nodes with limited connectivity and at most thousands of users. Most of the active links will be user links which are also normally end points (the only exception being when the user terminals are gateways to LANs or to legacy systems). The network topology is normally quite simple and the difference between the most efficient protocol and almost any heuristic routing protocol should be minor.

5.1 Routing for Mobile Users

We would first like to give a more detailed account of the operation of home and foreign agents than was provided earlier. Although we intend to use IPv6, to make Figure 14 simpler we resorted to IPv4 addressing because it is shorter. The concepts do not change.

This example both clarifies how mobile routing could be implemented in the mobile WI and prepares one for considering the issues of mobile multicasting with its complex address translation and routing issues. At this point we are only considering a simple unicast to a mobile user from outside the mobile WI AS. It is assumed that mobile user #1 (labeled ms#1) has moved from its home network (128.1.x.x) and is now temporarily attached to the subscriber network attached to airborne platform #1. ms #1 will always retain its home IP address. (We are using IPv4 class C addresses with subnetting only as an example. The entire mobile WI has a single net address.)

First, ms#1 registers with the foreign agent for the network (at 192.2.1.1) which represents the entire wireless WI to the outside world. (We will discuss registration in more detail elsewhere.) In turn, the FA informs user #1's home agent that user #1 is in the mobile WI and all external transactions need to be directed towards the FA #1.

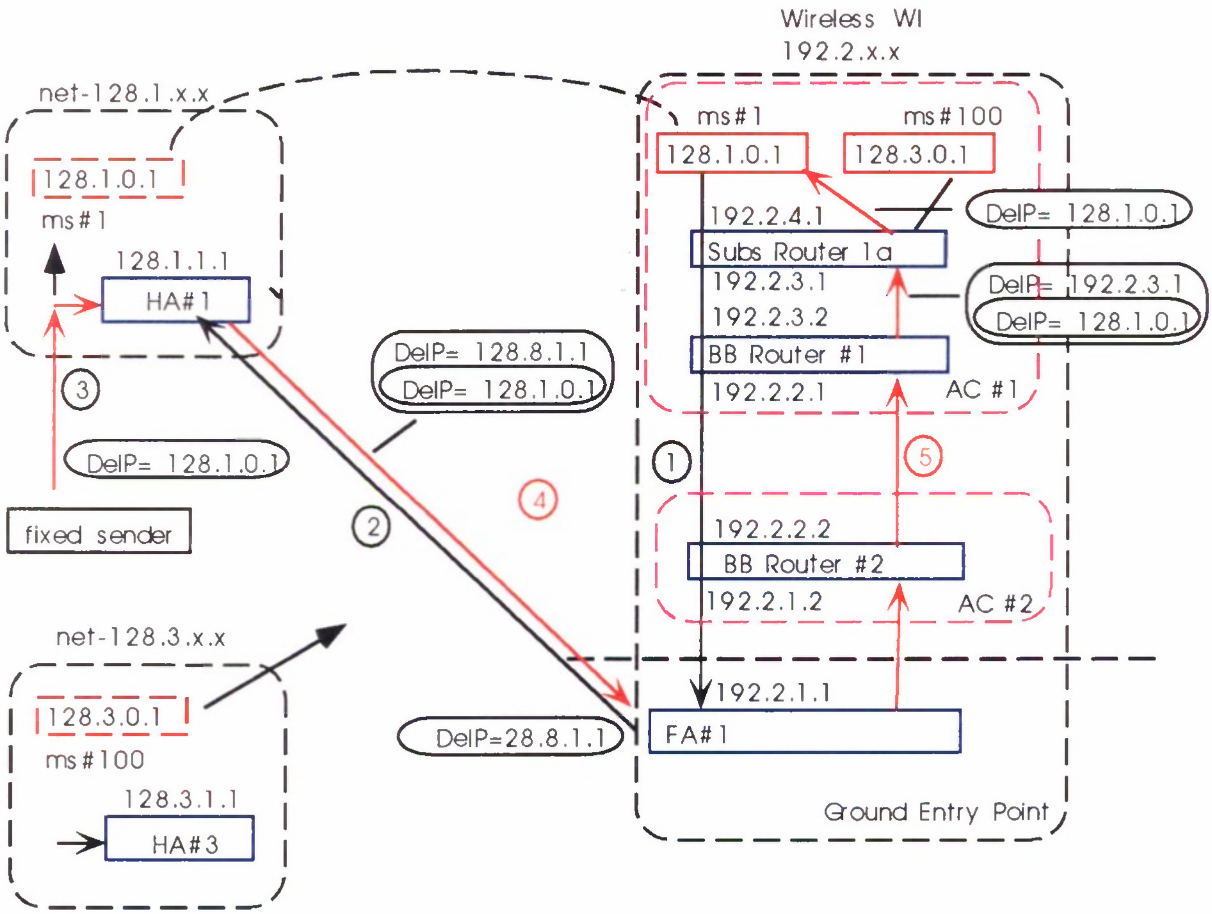
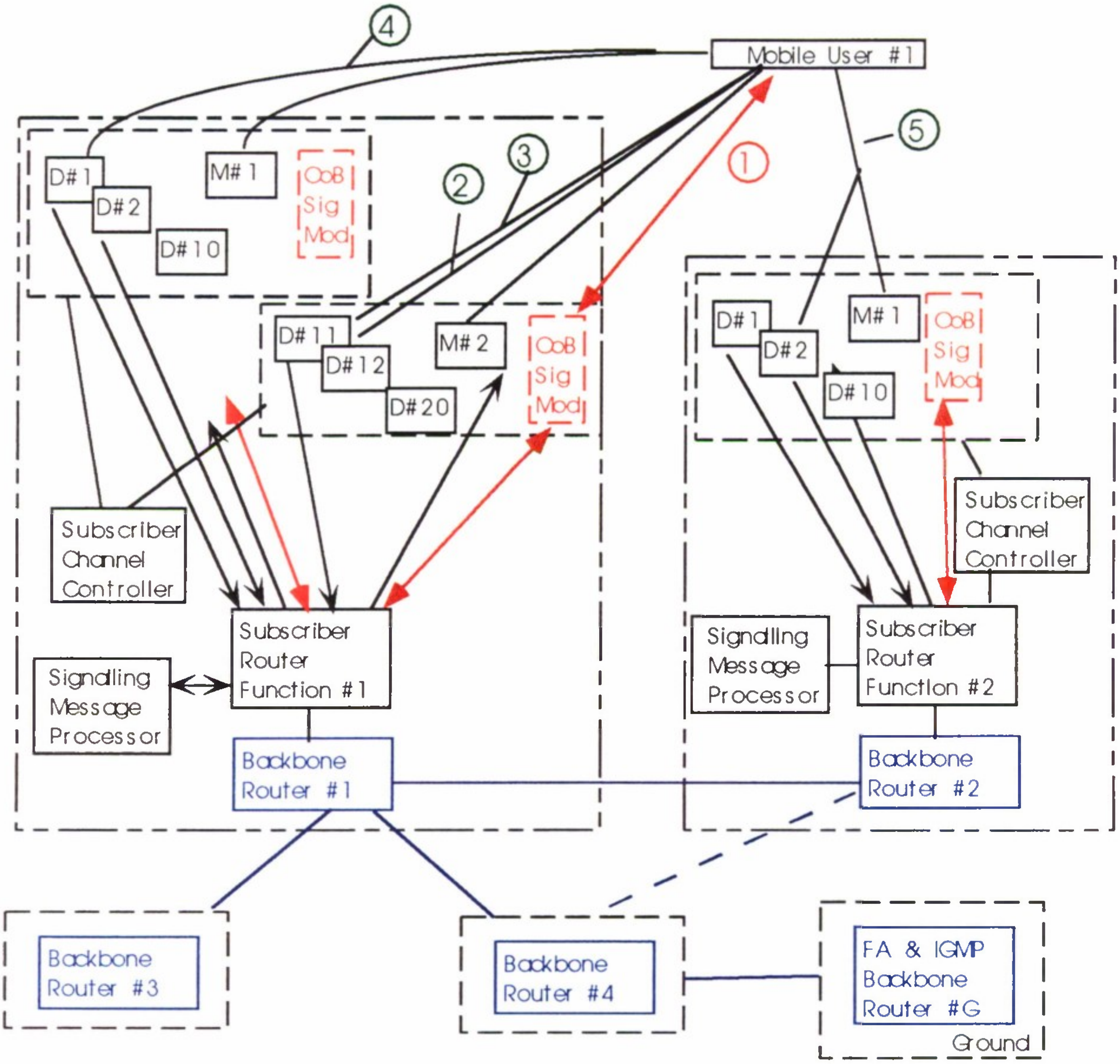


Figure 14. Example of mobile routing in WI.

Now assume that an external fixed user wishes to sends a datagram to ms#1 and uses as the destination address the home IP address of ms#1. This datagram will be directed to the home network. Since ms #1 is not there, the home agent router would capture the datagram, encapsulate the datagram with the FA IP address and route (tunnel is the common term) it accordingly. At the FA router (FA#1), the original encapsulation is removed and FA#1 could reencapsulate the datagram with the IP address of the subscriber router on airborne platform ap#1. It is assumed that as part of the registration process, this subscriber router has a map of the destination user IP to downlink channel and can decapsulate the packet and send it to the user who can filter on his home IP address. (Actually, there should be an equivalent of a MAC address filter used. This is an implementation detail.)

It is useful to consider mobile-user to mobile-user communications within the mobile WI. For this purpose, consider another user (ms#100) whose home base is served by Home Agent #3. We will make the assumption that this user moves to the mobile WI, registers, and then FA#1 informs HA#3 of the move so that external communications can be directly correctly. The registration process also was noted by the local subscriber router. Now let us assume that ms#1 wishes to communicate with ms#100. ms#1 only needs to include ms#100 home IP address as the destination address and the subscriber router will be able to handle the routing locally by searching for the destination address on its local active IP cache. This will result in the shortest route selection.

Let us look at the next level of mobility detail associated with reassignment of channels for any number of reasons. In Figure 15 we show the possibility of multiple downlink beams (only 2 are shown on airborne platform ap#1, but the extension to more beams is obvious). The figure represents four airborne platforms, all configured similarly (although the details are suppressed). We also have not shown the cross-link managers, but it is assumed that these managers have been used to establish the backbone links used by the routing algorithms.



4
Mobile User #1
5
3
2
1
D#1
D#2
D#10
M#1
OoB Sig Mod
D#11
D#12
D#20
M#2
OoB Sig Mod
D#1
D#2
D#10
M#1
OoB Sig Mod
Subscriber Channel Controller
Subscriber Router Function #1
Signalling Message Processor
Backbone Router #1
Subscriber Channel Controller
Signalling Message Processor
Subscriber Router Function #2
Backbone Router #2
Backbone Router #3
Backbone Router #4
FA & IGMP Backbone Router #G
Ground

Figure 15. End point mobility and rerouting.

When we earlier discussed mobility, it was with respect to a mobile user leaving his home network and attaching to the mobile WI. Here we want to discuss the situation where the mobile user is in the tactical theater and is either stationary or moves at walking or low vehicular speeds. The mobility considerations arise because the footprint of the airborne vehicle is moving, and the user (even if stationary) will lose his current coverage and (1) move under the coverage of another beam on the same airborne platform, (2) move under another beam of a different airborne platform, or (3) lose coverage completely. The latter is the worst situation and can only be ameliorated by rerouting an airborne platform to provide the coverage.

We need to clarify the roles of the subscriber router versus the backbone router. Although the functions of these two routers can be combined in a physical realization, it is easier to separate the functionality for tutorial purposes. Figure 15 illustrates the impact of a mobile user switching between different subscriber channels. We show 5 different cases.

1) The user is initially connected to the out-of-band signalling channel which is also used for call alerts and pages. The subscriber router will then route any information down on the associated downlink signalling channel. (This channel could be a low data rate TDM slot on the group's downlink broadcast or one could create a separate low rate downlink stream.)

2) When a service request is granted, the user switches from signalling assets to data channel assets. The subscriber channel controller coordinated this switch and by informing the subscriber router of this change, any downlink routing can be handled correctly. The backbone router has no knowledge of any such change.

3) There are situations where a user occupying a certain channel (up and down) may be switched from one data channel in a group to another in the same group. Again, only the subscriber router need be cognizant of this change.

4) If there are multiple subscriber beams on the airborne platform, then one routinely switches between data channels from group to group (beam to beam). Again, the beam handover process will be entirely confined within the subscriber system for that airborne platform.

5) Finally, if the user experiences a shift in beam coverage from one airborne platform to another, then we need to involve the backbone router as well. Only in this case is the user end point location updated in the backbone routers (including the all-important foreign agent router at the edge of the mobile WI). In all other cases, the end point location movement is handled by the subscriber router in conjunction with the subscriber channel controller.

Confining knowledge of end link "local" changes within the view of an airborne platform to the subscriber node can vastly reduce updates to routing tables as well as to multicast group tables.

The registration process is a key step in setting (and updating) the mobility information needed for routing (for both unicast and multicast). The log-on process is accomplished via the sharable (out-of-band) signalling channel associated with every group. The log-on messages are routed to the well-known address of the authentication server. Assuming the user is legitimate, the authentication center sends a message to the subscriber router which is reflected to the subscriber channel controller so that it is now conditioned to accept communication request messages from the user with the specified priority levels.

The channel controller can also send a message to its subscriber router which registers the user's IP address (and any other associated information) and could send messages to all the other subscriber routers via the backbone routers. A message generated by a mobile user will contain a destination IP address. If the local subscriber router does not recognize this IP address as being local to the aircraft, it could encapsulate the packet with the appropriate subscriber router IP address and send it out via the backbone router. The normal routing protocol will then deliver the packet to the proper subscriber router. (Of course, this is only one possible way to handle the routing determination.)

When the user coverage is shifted between airborne platforms a reregistration process takes place. This is necessary both to update all the routing tables and also to set the communication privilege levels that the user has. Perhaps this information could be exchanged in a channel controller to channel controller interaction rather than having to go to the authentication server. The former would certainly be faster and provide a better handover performance.

5.2 Routing Over the Backbone

The Mobile Networking Architecture Project (Monarch, Carnegie-Mellon University) has the goal to develop seamless wireless and mobile host networking. The scope of the work includes protocol design, implementation, performance evaluation, and validating acceptability at the user level. As such, the considerations span up to the application level layer in the conceptual ISO model. Their project is within the context of the Wireless Andrew project which builds on the wired network infrastructure at CMU. This fixed network consists mostly of interconnected Ethernets. For high speed wireless on campus, the transport is AT&T's WaveLAN, a DS 900 MHz radio operating in the ISM band. Off campus, CDPD is being employed. Monarch provides an implementation based on the basic IETF Mobile IP protocol and the more complex Mobile IP protocol with Route Optimization. The specifics of these mobile protocols are discussed elsewhere.

We find the most interesting aspect of Monarch is its relevance to the backbone mobility issues. Monarch's Johnson and Maltz define "an ad hoc network is a temporary network, operating without the aid of any established infrastructure or centralized administration." [This is certainly a characteristic of the WI backbone.] Any such network needs a routing protocol to dynamically find multihop paths between end point communicating entities. In the wireless environment, this routing must be sensitive to drastically changing propagation conditions and interference on any of its links. For a variety of reasons, the Monarch project team rejected both the conventional *distance vector* and *link state routing* schemes. They designed a *dynamic source routing* scheme.

In source routing, the sender of a packet determines the complete sequence of intervening nodes between itself and the destination point. The entire path is encapsulated in a header, and once released each intervening node knows where the entire packet is to be relayed next. In fixed networks, the routes can be determined statically or dynamically constructed. In mobile networks, only a dynamic source routing choice is possible. Over time a particular node participating in source routing builds up a database of possible destination nodes and how to get to that destination node. It does this by processing the source routing headers. Initially when it has no determined paths and whenever it does not know how to relay a packet, it invokes a *route discovery* protocol. In the route discovery, the sender transmits a *route request* packet to its neighbors, and these are relayed further (with the relay adding its address) until either one reaches the destination or to a host that knows how to get to the destination. The knowledgeable end point would then generate a *reply* packet which then could contain all the node addresses from end to end. There are a number of algorithms which eliminate loops and produce shortest paths. However, it is noted that the route discovery is only used as needed and there is no fixed periodic routing messages. The burden is that the intervening nodes need to store potentially all the routes to each end node. This is certainly an important issue of scalability but the Warfighter's Internet may not be of sufficient scale to make this a problem.

The problem still remains how to account for the situation when an active path is "broken" for any number of reasons. The source is tasked with the responsibility of monitoring the health of the end-to-end connection (this is part of the route maintenance function). How it receives its information is variable. It can use either active or passive acknowledgments at either the end-to-end or have information relayed by the individual link level acknowledgment process. In the latter case, any intervening host that senses an error has the responsibility of returning a route error packet to the sender of the failed packet so that this sender can construct an alternative path. In the Monarch project, the efficiency of their routing protocol appears both robust and efficient. (This appears to be a very attractive option for the Wireless Warfighter's Internet.)

Additionally, since the quality of different links in the path is variable and different connectivities are encountered in the same end-to-end point session, some mechanism is needed to reflect these changes in the using application. Hence, there needs to be some way in which lower layer protocols such as mobile IP needs to inform the upper layers of specific changes in the communications capabilities. The Monarch project intends to develop a new API and some extensions to mobile IP to provide notification services to mobile aware protocols and applications and to non-mobile hosts that the mobile host may be communicating with. One would need information on available BOW, delays, link error rates, etc., with respect to the mobile host's wireless path segments.

5.3 Multicast Routing in Mobile WI

Before considering multicasting let us review broadcasting. In flood broadcast, a router listens for a well-known broadcast address. When the router receives the broadcast, it relays the broadcast on all outgoing links with the exception of the link that the broadcast came in on. If another copy of the same message arrives, the router will drop it; this prevents loops. Note that there is no database of sources; everyone has the privilege of broadcasting and everyone will listen. There is also no surprise that a broadcast packet came in on the link that it arrived on. Every router executes broadcast. It is simple.

Let us now transition to the multicast situation. First, a user may be associated with many multicast groups so it must recognize a corresponding number of addresses. Belonging to a multicast group does not mean that multicast traffic will automatically be sent to you. Only if one is an active member will the message be explicitly sent to you. Thus, we have two lists to maintain for each multicast group: a list of all the members, and a sublist containing the active members. All routers that support multicasting must support a multicasting routing algorithm. However, the basic operation of the router with respect to incoming multicast packets is the same. If a multicast packet comes in on a specific link and that link is the expected arrival link given the packet source address, then the packet is replicated and sent out only on those outgoing links ultimately serving the active multicast group users. The intent is to minimize and confine overall traffic to as few links as possible. If the multicast packet arrives on an unexpected path (based on the multicast routing protocol), then this packet is dropped. The problem is that with mobility, packets from sources come from unexpected directions and may be dropped even if these are the legitimate routes.

As a result of the scale of the mobile WI and the fact that this network is an autonomous system, we can adopt some easy implementations that still are relatively efficient. In the multicast routing standards there is a proposal called Core Based Tree (CBT). The CBT multicast proposal uses a single tree for each multicast group. A specific router is designated as the core and a source of all multicasts. The source node sends the unicast datagram to this designated router and this router performs the multicast function.

The CBT concept seems "just right" as a starting basis for the mobile WI. The key concept is to use as the "core" router the same multicast-capable router that hosts both the foreign agent and the multicast manager. This multicast manager need only determine the minimum tree to get to all the active members. Fortunately, since it is also the foreign agent and the keeper of the active multicast lists, it has sufficient information to build the optimum distribution tree.

Let us assume a multicast group consisting of members (many mobile) entirely contained within the mobile WI. The multicast manager functions in conjunction with the active core tree generator to create the core tree with all the active members being in the tree. Let us further assume that member "a" wishes to send a multicast datagram. The process is to create a unicast message to the core route and then let the core route provide the multicast to all the members (actually to the subscriber routers serving the active members).

To simplify the routing, one could simply broadcast over the backbone and leave the multicast filtering function at the various subscriber routers. This topic is explored further in Appendix F.

Some of the open source code for mobile IP can be found at:

CMU - Mobile IP Daemon (for Linux)
http://www.ini.cmu.edu/^as26/mip/

SUNY (for LINUX)
http://anchor.cs.binghamton.edu/~mobileio/

Bibliography

M. Taylor, W. Waung, M. Banan, Internetwork Mobility - The CDPD Approach, Prentice Hall Series in Computer Networking and Distributed Systems, 1996.

M. Sreetharan and R. Kumar, Cellular Digital Packet Data, Artech House Publishers, 1996.

APPENDIX F
MULTICAST CONCEPTS FOR WARFIGHTER'S INTERNET

1. Introduction

Although it is an accepted theoretical fact that multicasting techniques can provide the most efficient use of network resources for a point-to-multipoint communication session, there are several extenuating circumstances that bear closer study in this real world application. Warfighter's Internet (WI) will operate in a dynamic, mobile environment, while the Transmission Control Protocol (TCP) and Internet Protocol (IP) that will presumably form the basis for this connectionless, packet-switched datagram network are inherently designed for a fixed station addressing paradigm. So it is not clear that we can achieve true multicast capability without some type of currently indeterminate TCP-IP enhancement.

An additional consideration is that WI has a unique, "skinny" topology, featuring a sparsely-connected airborne backbone network (BBN) with up to approximately 12 nodes, each with a dense cluster of as many as 1000 mobile communication devices (MCDs) on the ground arranged in a star configuration about the given airborne node. Even if we could circumvent the TCP-IP addressing limitation for mobile end users, it is not evident how much of an improvement multicast techniques within the BBN will actually provide for the particular network topology of WI. This appendix illuminates some of these issues by quantifying the relative network resource efficiency of multicast and other alternative suboptimal point-to-multipoint communication modes.

2. Illustrative Example

Consider a hypothetical multicast scenario in which a situation awareness update needs to be sent from an Enhanced Combat Operations Center (ECOC) through a WI airborne BBN to a designated group of MCDs on the ground. As noted above, the BBN is assumed to contain up to about 12 nodes, comprised of unmanned aerial vehicles (UAVs), Airborne Communication Nodes (ACNs), and other aerial platforms of convenience, connected via wireless T1 (i.e., 1.544 Mb/s) data communication links. The RF links between the airborne BBN and the ground nodes (including the ECOC) are assumed to have less bandwidth or capacity (e.g., 10–100 kb/s) than the airborne cross-links. The MCDs might be associated with small unit operations (SUO) or special operations forces (SOF) behind enemy lines, each of which is transiently affiliated with one of the airborne nodes.

To illustrate the WI communication concepts of interest, consider the specific BBN in Figure 1 consisting of a constellation of 7 airborne platforms in a topology analogous to a configuration of densely packed hexagonal cells in conventional terrestrial mobile telephony. A proposal was made at WI Study Team Meeting #5 to keep the cell orientation fixed relative to the ground in deference to the stationary addressing needs of the current Internet Protocol Version 4 (IPv4), and to have the airborne nodes and MCDs modify their cellular affiliations as they move around. We prefer to adopt a node-centered cell structure (i.e., one that is tied to the transient locations of each of the BBN nodes). While this shifts the burden of cellular affiliation to the MCDs, which must compensate for the motion of the airborne BBN nodes even if they themselves remain stationary, the node-centered structure more naturally reflects the logical connection between MCDs and their associated BBN nodes. In addition, requiring the BBN nodes to handoff MCDs would require each BBN node to know the ground footprints of all BBN nodes in real time, as well as the positions of the MCDs. In light of these considerations, the node-centered approach seems preferable.

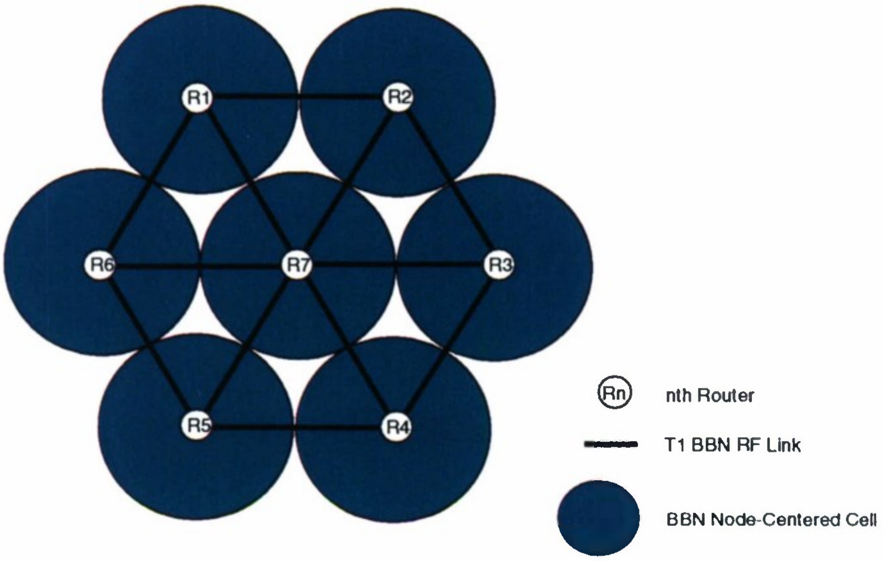


Figure 1. Cell topology and airborne backbone network (BBN).

Each airborne node is assumed to contain an RF transceiver with either an omnidirectional or a multiple-beam antenna (MBA), along with an intelligent router that has global knowledge of the entire dynamic BBN network topology for routing purposes. (We neglect issues of implementation and overhead associated with the distributed routing information in this paper; this is a relatively simple problem that has been solved before, and the required overhead should normally be small relative to the overall data throughput). We assume T1 cross-link connectivity only between adjacent aerial nodes rather than full BBN connectivity. The cells shown in Figure 1 can be thought of as plan view footprints on the ground representing the RF transceiver range for reliable communications between the aerial nodes and the MCDs on the ground. For example, if each BBN node can provide coverage over a 100-mile diameter area on the ground, the tightly packed constellation of 7 airborne platforms shown in Figure 1 would cover a 300-mile diameter ground swath with essentially no communication gaps. Ideally, these 7 airborne platforms would move in concert on a coordinated race course so that the relative orientation of the BBN is more or less maintained, and the entire pattern moves accordingly along the ground. Of course, in the real world, with aircraft other than UAVs used as BBN nodes of convenience and with nodes entering and leaving the BBN during the normal dynamics of tactical warfare, this kind of coordinated node motion and resulting uniform ground coverage will not generally be possible; furthermore, if there are only a few isolated clusters of SUO or SOF mobile subscribers to contend with, uniform coverage may not even be necessary.

Because of the mobility of the ground end users, and the requirement for fixed IP addresses, we will need to implement a home agent (HA)–foreign agent (FA) collaborative system (e.g., [1]) with FAs resident in each airborne node. As the ground and airborne nodes move around, an automatic registration scheme will have to continually affiliate each MCD with its nearest BBN node in a manner transparent to the users. Because of this constant change in network topology, the configuration shown in the attached slides should be regarded as a snapshot in time. Unicast routing can be accomplished by using the destination's home address in the IP header and having the HA forward messages intended for a given mobile host to its current FA; however, if the source and destination node are both MCDs affiliated with the same FA in a given airborne node, that FA will route the message directly from the source MCD to the affiliated airborne node and back down to the destination MCD, which avoids the inefficiency associated with a triangular route through the destination's HA.

The BBN configuration of Figure 1 is extended to the entire WI structure in the figures that follow. As noted earlier, the topology of interest is characterized by a "skinny" BBN with a relatively small number of airborne nodes, each of which can have a relatively large number of affiliated MCDs arranged in a star topology. The multicast subset in this example contains only 10 of these MCDs. Unicast, broadcast, and true multicast communication schemes are compared by following the transmission of a representative single packet of data from the ECOC to the 10 end users in the multicast group based on the number of times this packet must be replicated throughout the network, and the corresponding burden this places on the network resources, without distinguishing between the high-capacity airborne cross-links and the lower-capacity uplinks and downlinks.

Since the network connectivity is based on a wireless RF environment rather than a hard-wired implementation, we also need to address the multiple-access (MA) format. We anticipate that spread-spectrum radios will be used to provide some measure of transmission security (TRANSEC), including anti-jam (AJ) and low probability of detection and interception (LPD/LPI). This lends itself naturally to some form of spread-spectrum multiple-access (SSMA) scheme[1] which allows adjacent cells to reuse the available time-bandwidth resources.

[1] The alternative term, code-division multiple-access (CDMA), is often used in place of SSMA. This label is derived from the common (and inaccurate) regard for the pseudonoise (PN) sequence used to achieve the SS modulation as a "code" rather than a pseudo-random sequence, and is often assumed (also incorrectly) to be restricted to direct-sequence (DS) rather than frequency-hopped (FH) SS modulation.

There are generally two approaches to SSMA in a network context: source-oriented (or transmitter-oriented) and destination-oriented (or receiver-oriented) signalling. Very simply, in source-oriented SSMA, each BBN node transmits to all other connected nodes using a common pseudo-random SS pattern, and the burden is transferred to the destination nodes to demodulate the SS sequence appropriately. Destination-oriented signalling is the reverse arrangement in which, for example, the source node containing router R7 in Figure 1 must transmit the same packet using the 6 distinct SSMA patterns associated with each of the adjacent destination nodes if a broadcast operation is desired.

For the special combination of source-oriented SSMA and an omnidirectional transmitting antenna, a single RF transmission using the SSMA pattern associated with the source transmitter will service all of the intended receivers, provided that the medium access control (MAC) enables it. In all other cases (i.e., source-oriented SSMA with directional antennas, or destination-oriented SSMA with either type of transmitting antenna), a separate RF transmission is needed to reach all of the desired adjacent nodes. So source-oriented SSMA with omnidirectional transmitting antennas results in a much more efficient use of the RF network resources. For simplicity, the term "source-oriented SSMA" below will be used to succinctly connote this special case with omnidirectional antennas while "destination-oriented SSMA" will imply the less efficient RF signalling modes.

3. Comparison of Signalling Schemes

Consider first the destination-oriented SSMA approach. Figure 2 shows a standard unicast mode in which the ECOC must generate 10 identical data packets that differ only in the IP destination addresses within each header, and these must be transmitted via the ground-to-air data communications uplink to the BBN node associated with router R1. R1 in turn must forward 4 of these packets to the node identified by R2 using the SSMA pattern unique to that destination, and the other 6 packets to R7 (this is clearly not a unique route) with its particular SSMA pattern. R7 then retransmits 3 of these packets to R5 and 2 to R4. Then R2, R4, and R5 forward the packets to their intended MCD destinations. (We shall later see that Figure 2 is also valid for source-oriented SSMA unicast signalling.)

Assuming no retransmissions are needed, we see that this unicast example requires the combined transmission of 35 identical data packets (albeit with different IP destination addresses in their headers) over various RF link connections in the network. Since the example of Figure 2 uses unicast transmissions, any packets destined for nodes which are not attached to their HAs will have to be forwarded to the nodes' FAs for delivery, consuming additional network resources; no such packets are shown in the figure. Although we've illustrated the BBN links by a thick line reminiscent of a wireline network, it should be remembered that each node uses a wireless RF radio. So, for example, when R1 transmits to R2 and R7, if an omnidirectional antenna is used, node R6 will also receive this transmission, although it will ignore it after it looks at the destination address in the header.

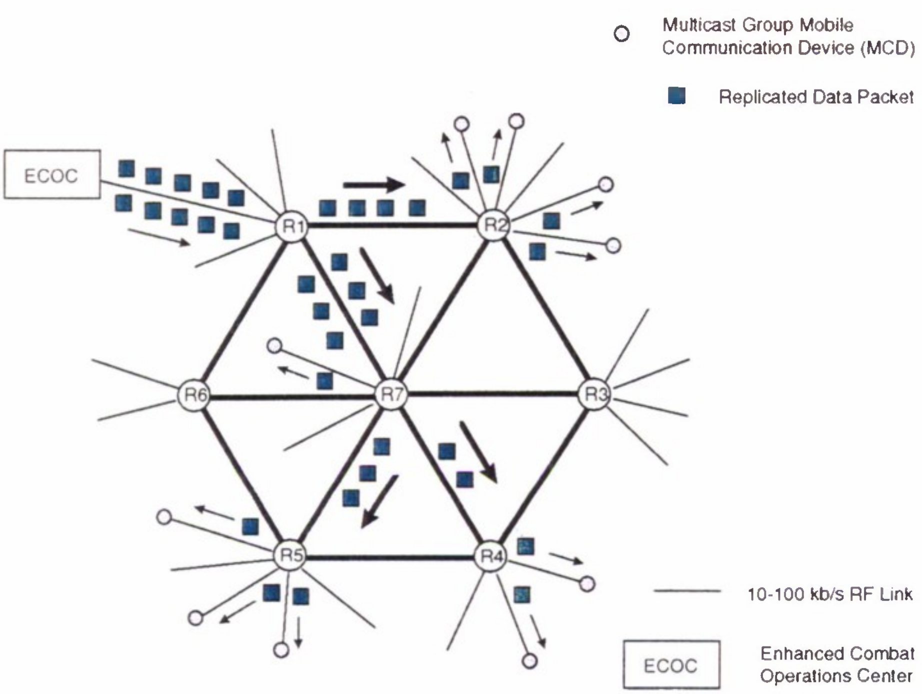

Multicast Group Mobile Communication Device (MCD)
Replicated Data Packet
ECOC
R1
R2
R3
R4
R5
R6
R7
10-100 kb/s RF Link
ECOC
Enhanced Combat Operations Center

Figure 2. Unicast mode designation- or source-oriented SSMA (35 replicated packets sent).

By comparison with unicast signalling, the destination-oriented SSMA broadcast flooding mode in Figure 3 transmits a single identical packet over each link in the network. Although care is taken to eliminate duplicate packet transmission over any link, flooding is still very inefficient since duplicate packets can still arrive at a given node (for example, R7 receives 3 identical packets from R1, R2, and R6). This scheme requires the transmission of 39 replicated packets, which is even less efficient than unicast signalling.

A more efficient destination-oriented broadcast scheme is achieved when we use intelligent routers in the BBN to eliminate the reception of duplicate packets. We assume that the destination field of the packet transmitted from the ECOC contains a multicast address in the style of IPv6 [2, 3]. In the so-called "intelligent broadcast" scheme of Figure 4, the ECOC sends out a packet to R1, which replicates it three times so that identical packets are forwarded to R2, R6, and R7 using three different SSMA patterns. Because of the particularly convenient topology of the BBN with R7 as a central hub, R7 then replicates the packet three more times and sends it on to the rest of the BBN nodes, R3, R4, and R5. Finally, each BBN node broadcasts the packet using the appropriate destination-oriented SSMA sequence for all of its active MCDs on the ground. This ultimately requires the combined transmission of 33 replicated data packets over the network. The difference between intelligent broadcast and unicast is that each packet is now only sent once to each node; however, there is still inefficiency in this signalling scheme because the identical packet is sent to every node in the network instead of just that subset necessary to reach the desired mobile multicast group.

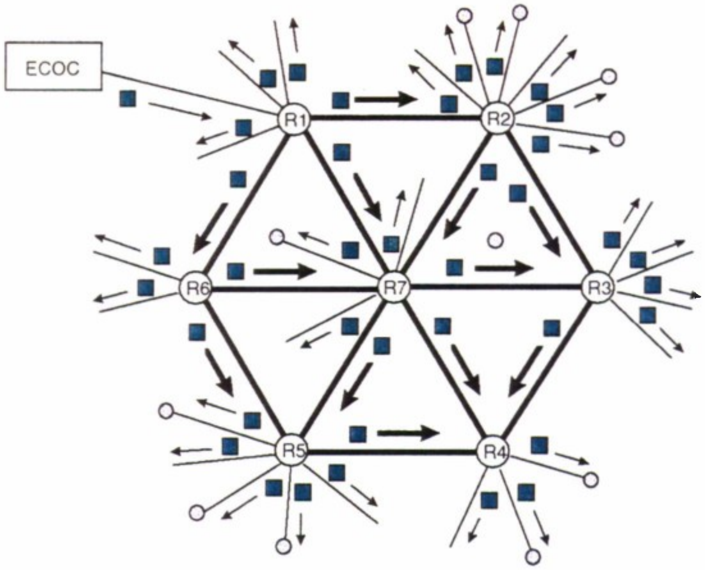


Figure 3. Broadcast flooding mode destination-oriented SSMA (39 replicated packets sent).

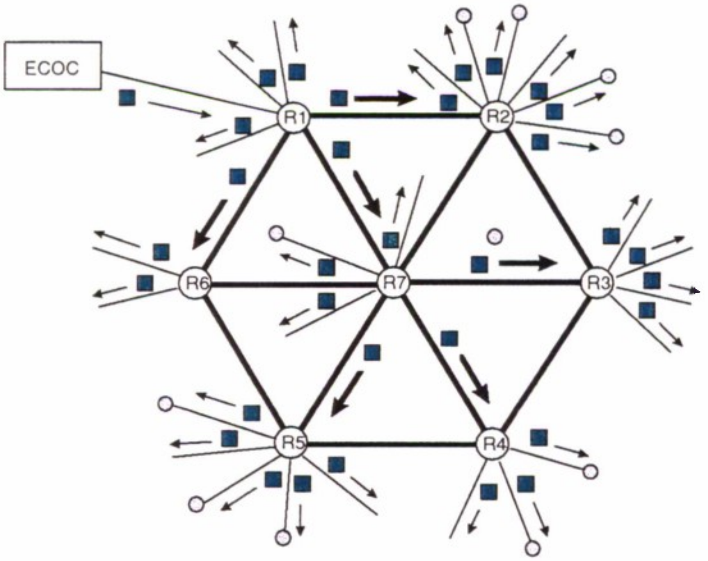


Figure 4. Intelligent broadcast mode destination-oriented SSMA (33 replicated packets sent).

In Figure 5, we demonstrate the savings in overall communication network capacity afforded by a true multicast scheme, without regard for whether such a paradigm can be effectively implemented under the limitations of currently available protocols. In fact, the current IPv4 and the evolving next generation IPv6 protocols will not provide a reliable multicast capability in a mobile environment without some additional enhancements or extensions. Nonetheless, disregarding the feasibility issue for the moment, it is assumed that the intelligent routers at the intermediate nodes can determine on which outgoing branches to replicate the transmitted packet in order to efficiently reach all of the members of the intended multicast group.

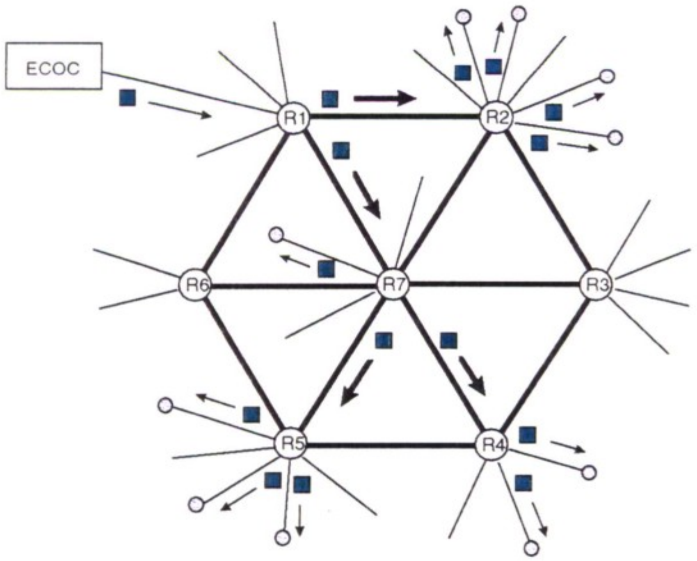


Figure 5. Multicast mode destination-oriented SSMA (15 replicated packets sent).

For example, R1 sends identical destination-oriented SSMA packets to R2 and R7, but not to R6. Similarly, R7 forwards the packet to R4 and R5, but not to R3. Finally, the 4 BBN nodes with MCDs in the multicast group only send the packet on to those particular 10 end users. It is shown that under this multicast scheme, the total number of replicated packets drops to 15, a reduction to 43% of the overall network capacity consumed by the original unicast approach.

Because of the shortcomings of IPv4 and IPv6, without some appropriate booster such as the one the SCPS-TP folks have begun to look at, these multicast paradigms cannot be implemented yet in a mobile environment. For this reason, we looked at some hybrid unicast-broadcast alternatives; i.e., we partitioned the complete WI network by separating out the ECOC and the airborne BBN from the air-to-ground data communication links and the ground mobile hosts and examined pseudo-multicast schemes in which the ECOC-BBN segment used either a local unicast or broadcast model followed by unicast or broadcast transmission from the FAs in the BBN to the end users. We still need to carefully examine the addressing issues to see whether the Class D group multicasting capabilities provided by IPv6, or tunneling/encapsulation techniques using an outer IP header addressed to the appropriate FA and an inner IP address with the final end destination, will permit these hybrid schemes to work.

With this disclaimer, we next looked at three distinctive destination-oriented SSMA signalling approaches shown in Figures 6–8: "segmented unicast (ECOC-BBN)-broadcast (BBN-mobile ground users)," "segmented intelligent broadcast-broadcast," and "segmented intelligent broadcast-unicast." The total number of replicated packets for these three schemes and the assumed network topology is 27, 24, and 17, respectively.

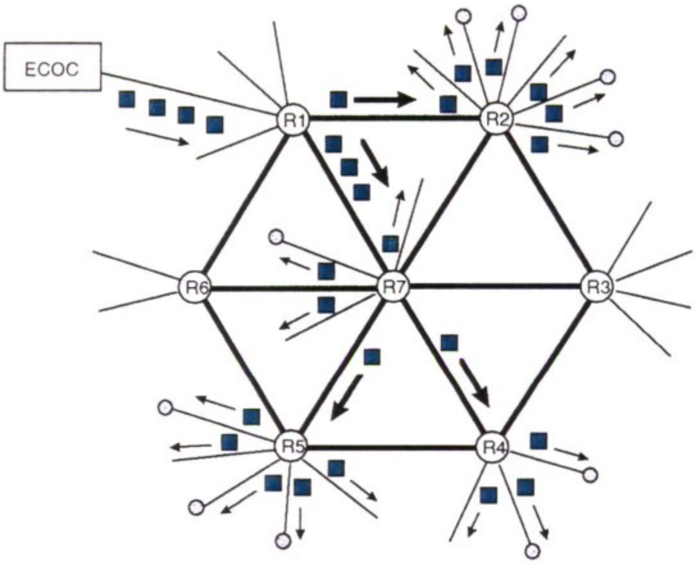


Figure 6. Segmented unicast-broadcast mode destination-oriented SSMA (27 replicated packets sent).

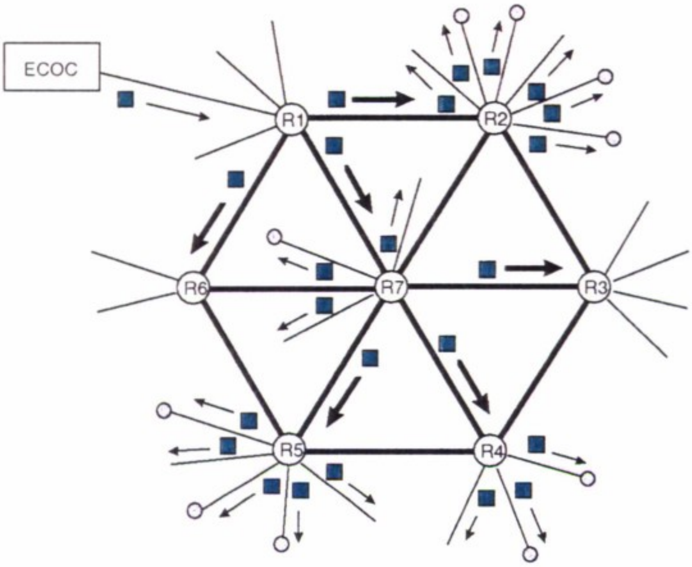


Figure 7. Segmented intelligent broadcast-broadcast mode destination-oriented SSMA (24 replicated packets sent).

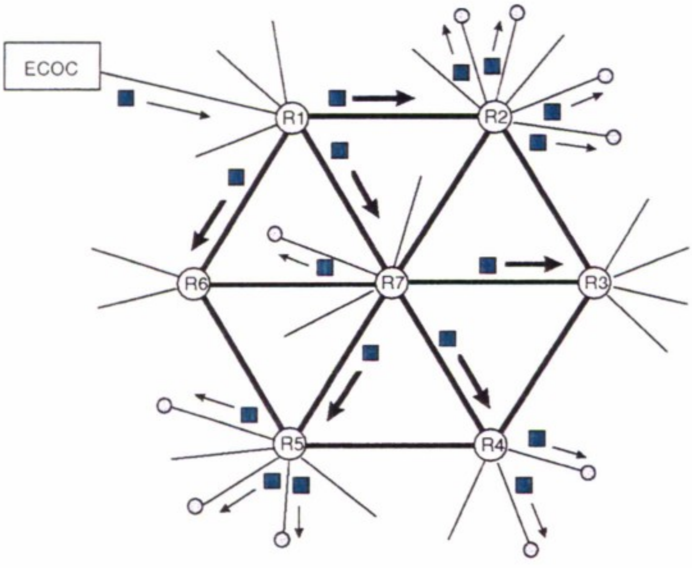


Figure 8. Segmented intelligent broadcast-unicast mode destination-oriented SSMA (17 replicated packets sent).

The savings in network communication capacity become more dramatic when we adopt the source-oriented SSMA signalling approach (with omnidirectional transmitting antennas in the BBN), except for the unicast mode which requires the same number of replicated packets as the destination-oriented SSMA scheme in Figure 2 because of the different IP destination addresses in the packet headers. For example, consider the source-oriented SSMA end-to-end intelligent broadcast mode in Figure 9. The ECOC sends a packet on its uplink to R1, which in turn broadcasts it simultaneously to R2, R6 and R7 using its unique SSMA pattern. A single RF transmission from R7 employing its SSMA sequence then forwards this packet to R3, R4, and R5. Finally, the 7 BBN nodes broadcast to all active mobile subscribers using their associated downlink SSMA signalling schemes. The net result is that the end-to-end broadcast is accomplished with only 10 replicated packet transmissions, which is more efficient than the destination-oriented SSMA multicast mode of Figure 5.

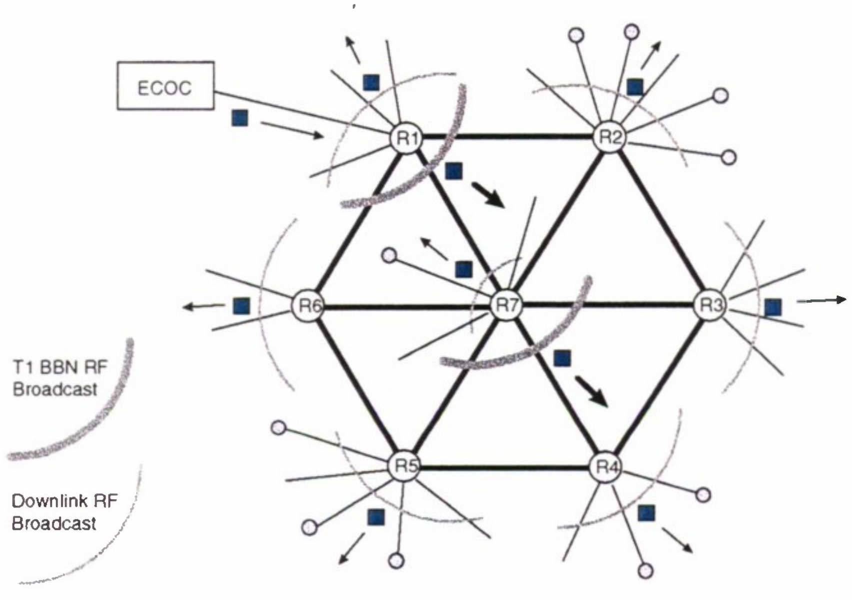


Figure 9. Intelligent broadcast mode source-oriented SSMA (10 replicated packets sent).

This is further reduced to only 7 RF packet transmissions (16% of the unicast requirement) with source-oriented SSMA multicast signalling as illustrated in Figure 10. It should be noted that this scheme has all of the IP addressing problems associated with destination-oriented SSMA multicast signalling. However, it can be shown that the segmented intelligent broadcast-broadcast scheme in conjunction with source-oriented SSMA signalling (also shown in Figure 10) provides the same performance as multicasting for the given topology without these addressing issues; this is the approach we propose for WI.

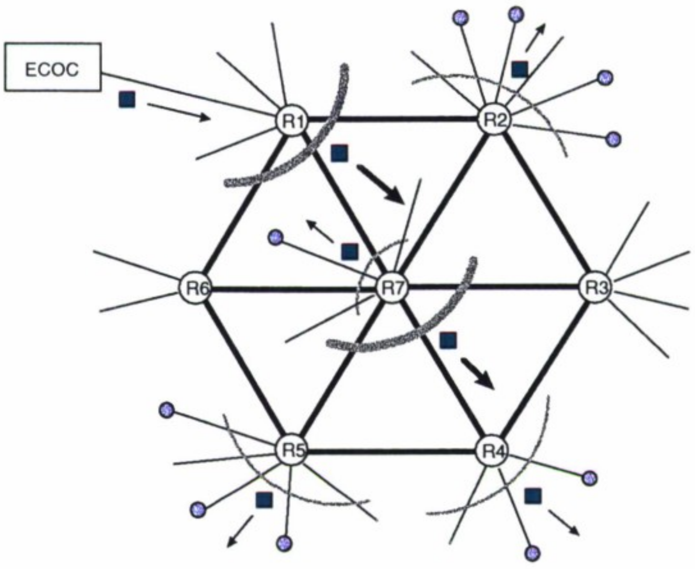


Figure 10. Multicast or segmented intelligent broadcast-broadcast mode source-oriented SSMA (7 replicated packets sent).

(Incidentally, although we do not explicitly illustrate the source-oriented SSMA versions of the segmented unicast-broadcast and broadcast-unicast schemes, it can be shown that they required 14 and 13 packet transmissions, respectively.)

When assessing these schemes, it is important to consider not only their bandwidth efficiency but also the overhead required to implement them. For example, while the segmented unicast-broadcast method saves some network bandwidth relative to full broadcast, hidden in it is the requirement that the ECOC know the set of routers supporting the current multicast set. If instead the ECOC transmits the packets to the MCD's HAs, some packets will almost surely have to be forwarded to their FAs for delivery. Contrast this with the segmented intelligent broadcast-broadcast scheme of Figure 7, in which case the ECOC does not need to know the locations of the MCDs, and the HA-FA aspect can be bypassed.

We now address whether or not the segmented approaches are reasonable using the multicast capabilities of IPv6. In short, IPv6's multicast support is sufficient to accommodate the segmented intelligent broadcast-broadcast approaches of Figures 7 and 10. By using a combination of host registration, whereby an MCD transmits its multicast group memberships on changing routers and the IPv6 Internet group management protocol (IGMP), whereby routers periodically poll for the group memberships of their attached hosts, the routers can track the multicast groups that they need to support. In particular, the registration process will serve to immediately notify a router of the multicast groups required by a new host, while IGMP querying will allow routers to determine when particular groups no longer need to be supported without requiring explicit de-registration of departing hosts.

Under IPv6, it is the multicast routers' job to ensure that multicast packets reach their destination networks (subject to the "time-to-live" parameter). To this end each BBN node will use its information of the dynamic BBN topology to maintain efficient broadcast trees. The source nodes for all multicast transmissions will be responsible for including in the hop-by-hop options of each outgoing multicast packet any information needed to facilitate this efficient broadcast. This information might include, for example, the address of the BBN router that serves the source as well as a time stamp.

As for the additional desire for reliability, we can envision a NACK scheme in which each FA-router stores the group multicast packets in a queue and performs a local retransmission to all of its affiliated mobile hosts if any of their multicast group subscribers did not correctly receive the message. This avoids the additional network capacity burden on the BBN of having the ECOC perform an end-to-end retransmission. Some explicit signalling may be included to inform senders of multicast hosts which persistently fail to acknowledge packets.

References

1. George Xylomenos and George C. Polyzos, "IP Multicast for Mobile Hosts," in IEEE Communications Magazine, January 1997, pp. 54–58.

2. William Stallings, "IPv6: The New Internet Protocol," in IEEE Communications Magazine, July 1996, pp. 96–108.

3. S. Deering and R. Hinden, Editors, "Internet Protocol, 6 (IPv6) Specification," RFC 1883, Xerox PARC, Ipsilon Networks, December 1995.

4. S. Deering, "Host Extensions for IP multicasting," STD 5, RFC 1112, Stanford University, August 1989.

APPENDIX G

FREQUENCY MANAGEMENT AND COORDINATION

Several recent discussions have persisted related to appropriate frequency selection and coordination for the Warfighter's Internet architecture development. The following subsections provide a listing of various potential frequency bands and the other users or potential users that the Warfighter's Internet would be competing with for the bandwidth. These listings were obtained from several sources that include several frequency spectrum and management guides and books, the U.S. Table of Frequency Allocations, as well as discussions with JPL's Frequency Manager, Twenty Nine Palms Frequency Manager, and Fort Irwin's Frequency Manager.

225–328.6 MHz and 335.4–399.9 MHz

Current Uses

This is typically a government military frequency band used for both fixed and mobile purposes. This band is principally used by military aircraft; tactical and training communications; satellite communications linking the activities of ground, air, surface, and subsurface mobile users; and rocket test and telemetry. On the commercial side of applications, car alarm and garage door opener manufacturers utilize this band as well for their handheld remote controls.

Significant systems include air-ground-air radios capable of hopping across many individual channels distributed across the 225–400 MHz range and military mobile-satellite systems with downlinks at 240–270 MHz and uplinks distributed across 235–322 and 335.4–399.9 MHz.

NASA also uses four of these frequencies in this military band: 243, 259.7, 279, and 243 MHz for Space Shuttle spacesuit-orbiter, orbiter-ground and emergency crew return communications. NASA is in the process of moving these frequencies to the 400 MHz range where it will have more flexibility to use them.

Also operating in 244–317 MHz is FLTSATCOM, the Navy's Fleet Satellite Communication System connecting aircraft, ships, submarines, ground stations, and command elements. The Air Force AFSATCOM system and the Presidential Command Network have channels on the FLTSATCOM satellites as well. The Navy is currently replacing FLTSATCOM with a UHF Follow-On (UFO) satellite system. This new system will operate in the same frequency band, but will be much more spectrally efficient. MILSTAR also has uplinks and downlinks in this band to keep compatible with the older satellite systems.

Future Outlook

The FCC has called upon the federal sector to make spectrum available in this band to the private sector, noting that other nations have found it appropriate to use portions of this band for nonmilitary purposes. Discussions between the FCC and NTIA are underway for use of 312–315 and 387–390 MHz for mobile satellite communications. This band is currently and will continue to be a highly congested one.

406.1–420 MHz

Current Uses

The 406.1–410 and 410–420 MHz bands are changing from supporting mostly fixed government to a land mobile government band. Key uses in these bands include links to connect mobile networks as well as for weather, hydrology, seismic sensors, law enforcement and protection of the President and other government personnel and foreign dignitaries, flood and wind-shear warning and telemetry for electric power transmission.

The U.S. Air Force, Army, and Navy uses include radar, paging and range control, disaster preparedness, training, shipboard operations, and security. The 410 - 420 MHz band is exclusively reserved for the Federal Government. Radio astronomy is the exclusive non-government user in the 406.1–410 MHz portion. A private corporation, FEDSMR, operates trunked systems for the Federal Government in this spectrum in a lot of the major U.S. cities.

Future Outlook

Trends in this spectrum include migration of the fixed systems to the 932–935 and 941–944 MHz band, and increased use of multiagency trunking systems. This band is currently and will continue to be a highly congested one.

1350–1400 MHz

Current Uses

This band plays several roles in military communications, aerospace, Earth Science, and radio astronomy. The 1390–1400 MHz segment has been proposed for gradual reallocation to the private sector for as-yet-to-be-determined uses. The proposal would remove most federal operations from this segment by January 1999. The FCC has claimed that allocating only 10 MHz of this band would severely limit potential uses. Accordingly, it has asked for more of this band to be made available for nongovernment use.

A little known function of the Navstar Global Positioning System (GPS) uses a 5 MHz channel centered on 1381.05 MHz to transmit an alerting signal to fixed and mobile receivers as part of the Nuclear Detonation Detection System (NUDET). The U.S. Air Force is the largest government user of 1350–1400 MHz, for high-power, long-range, surveillance radars. The Federal Aviation Administration (FAA) and the Department of Defense (DoD) operate the nationwide Joint Surveillance System (JSS) radar network around the U.S. perimeter in 1240–1370 MHz for civilian aeronautical radionavigation and for air defense, fleet defense, and drug interdiction. January 1994 saw the launch of a new U.S. Air Force system in this band, the Range Joint Program Office GPS Data Link, used for rebroadcasting position information of high-velocity manned and unmanned aircraft during test and training events. This system is used at all major military aircraft and missile test centers. It is used in flight testing of aircraft such as the B-1 and B-2.

Future Outlook

Current use of this band by the military and civilian government is substantial, and a significant amount of resources have recently been invested in this band making it highly unlikely for future military expansion.

1429–1435 MHz

Current Uses

This frequency band is used for telecommand of missiles, experimental testing, tactical and training, light route radio relay, and radar cross-section measurement. The U.S. Navy uses this band for testing of shipboard electronics, command of remotely-piloted vehicles, and research. Aerostat balloons used to detect low-flying aircraft suspected of carrying drugs use this band among others for data and/or voice links. Nongovernment uses include antenna testing, automation, and manufacturing.

Future Outlook

This particular band is not as over-subscribed as some of those previously mentioned. This band probably has some strong potential for future military expansion with proper coordination and preparation.

1435–1530 MHz

Current Uses

This frequency band is considered extremely important for U.S. aeronautical test telemetry. It is used for flight testing of manned and unmanned aircraft, missiles, and space vehicles, and associated communications such as range safety, chase aircraft, and weather data. In fact, this band is so congested that new systems are being moved to flight test spectrum at 2360–2390 MHz.

Future Outlook

The FCC has proposed to modify the allocations within this band to conform to changes made to the international table of allocations at the 1992 World Administrative Radio Conference (WARC-92). The international changes were intended to supplement the spectrum available to the mobile satellite service. The changes would create a 1435–1525 MHz band limited to flight testing on a primary basis, and a 1525–1530 MHz band allocated to mobile satellite downlinks on a primary basis and flight testing on a secondary basis. This band is not likely to be made available for future military expansion like for Warfighter's Internet.

1710–1850 MHz

Current Uses

This is an exclusive Federal Government frequency band. At this time it contains more than $10 billion worth of ground and space system assets. The U.S. Air Force Satellite Control Network (AFSCN) and the Space-Ground Link Subsystem (SGLS) use the 1761–1842 MHz segment of this band (uplinks) and 2200–2290 MHz (downlinks) to control defense and research satellites and Space Shuttle functions, from primary stations in Guam, Hawaii, New Hampshire, Colorado, and California, and from transportable stations that provide additional coverage for launch and orbit operations. The Defense Meteorological Satellite Program (DMSP) satellites use these bands to provide real-time weather data to U.S. Air Force, U.S. Navy, and U.S. Marine Corps tactical ground stations and ships.

Other military uses of this band include tactical and transportable radio relay systems for ground mobile forces; air combat training systems that compute and transmit altitude, velocity, and weapons status in simulations; television from aircraft-mounted cameras for remote piloting and monitoring of civil disturbances; weapon and target scoring systems; and robotic telecommand for hazardous waste cleanup. The balloons used to detect low-flying aircraft suspected of drug smuggling use voice and data links in this band as well.

This is a major government band for medium-capacity point-to-point microwave links. Federal fixed service users include the Tennessee Valley Authority for control and sensing in large electric utility operations, the FAA for low data rate radio links, and the National Forest and Park Management for communications in remote areas. The Army Corps of Engineers operates a backbone communications network in this band for remote controlled hydropower stations, support for the Federal Emergency Management Agency (FEMA), and flood control and maintenance communications for inland waterways, harbors, locks, and dams.

Future Outlook

The 1710–1755 MHz segment has been proposed to be cleared of some federal users by January 2004, for eventual re-allocation to the private sector for as-yet-to-be-determined users. Emergency, military, and federal power operations will continue to operate in this segment even after re-allocation.

Licensees and manufacturers in the Personal Communications Services (PCS) look forward to re-locating some non-government fixed microwave links from the PCS bands into the 1710–1850 MHz band. The FCC has asked that more than 1710–1755 MHz be re-allocated to the private sector, and that it be done before 2004. The outlook for this band for future military expansion is highly unlikely due to its already congested nature.

2200–2290 MHz

Current Uses

For national security reasons, this band is available to the Federal Government only. It supports tracking, telemetry, and control for federal space programs, and numerous conventional fixed microwave systems and military aviation, tactical and training operations. Key uses include downlinks for both the NASA Tracking and Data Relay Satellite System (TDRSS) and the U.S. Air Force SGLS. This band is used intensively for space operations and research, with agreements for cross-support amongst space agencies. It is the main downlink band that NASA uses to receive scientific and engineering data from spacecraft. Some of the scientific spacecraft in this band include the International Ultraviolet Explorer, the Dynamic Explorer System, the Advanced Magnetospheric Particle Experiment/Charge Composition Explorer, the Earth Radiation Budget Satellite, the Cosmic Background Explorer, the Gamma Ray Observatory, and the Roentgen Satellite.

This band also provides terrestrial telemetry in support of nuclear, airborne weapons and flight testing, intership data relay, missile range operations, and for radar-equipped tethered balloons. This is also one of the main bands for air-to-ground video transmission on military test ranges.

Future Outlook

No new or future plans are evident for this particular frequency band. Proper coordination for future military expansion would need to be accomplished not only within DoD but also NASA.

2290–2300 MHz

Current Uses

This is a primary NASA Deep Space Network (DSN) band for telemetering data from probes outside Earth's orbit. DSN uplinks are at 2110–2120 MHz. Spacecraft using this spectrum include the Pioneer and Voyager interplanetary probes, the Magellan mission to Venus, the Galileo mission to Jupiter, the Cassini mission to Saturn, and Ulysses, which is examining the Sun. This band is used for Very Long Baseline Interferometry (VLBI) observations using multiple radio telescopes. The band also is important for measuring the polarization of space radio emissions.

Future Outlook

NASA has invested multiple billions of dollars into its DSN and would be highly reluctant to relinquish any of this band for other uses. The DSN receives extremely low signal levels from planetary spacecraft, at distances where the signal transmit time is often measured in hours. This type of communications requires very large antennas (up to 70 m) and ultra-low noise amplifiers that are cryogenically cooled. NASA is extremely sensitive to monitoring the commercial sector's use of the adjacent frequency bands for this reason, much less relinquish part of its own band.

4400–4500 MHz

Current Uses

This frequency band is allocated for federal use for fixed and mobile operations on a shared primary basis. The major users of this band are the DoD, Department of Energy (DoE), and the Treasury. Links in this band also connect major U.S. Army headquarters. It is the only band used by the U.S. military for tropospheric scatter communications. Other military uses are video and radar bomb scoring, air-to-ground video, air combat maneuvering instrumentation, drone aircraft control, microwave links associated with land mobile radio systems, and tactical relays. The U.S. Navy uses of this band include data transfer between ships and helicopters, control of drone aircraft, and developmental aeronautical systems. The broadcast television industry has identified this band as desirable for auxiliary links needed to carry High Definition Television (HDTV) between broadcast sites as well.

Future Outlook

No new rulings on this frequency band are expected in the near future, although additional recent pressure from the television industry related to digital television and HDTV broadcast could change that fact. Should this band stay exclusively federal government use, it has some strong potential for supporting a system configuration like that for Warfighter's Internet.

7.25–7.75 GHz and 7.90–8.40 GHz

Current Uses

The Defense Satellite Communications System (DSCS) downlink is at 7.25–7.75 GHz. The uplink is at 7.90–8.40 GHz. The geostationary DSCS II and III satellites provide the DoD, State, and other agencies multiple-beam, jam resistant voice and data communications. Also in this band are the two NATO 4 satellites that provide voice and data communications between members of the North Atlantic Treaty Organization, with command and control by the U.S. Air Force. These satellites are scheduled to reach the end of their design lives by the end of the decade.

Future Outlook

With proper coordination with the DSCS user community, the band would seem highly attractive for future military expansion. This band is not overly congested at this time.

13.4–14.0 GHz

Current Uses

NASA's TDRSS downlink is at 13.4–14.05 GHz. The Space Shuttle carries a rendezvous radar for satellite retrieval in this band. This is also one of the bands police radars use.

Future Outlook

With proper coordination with the NASA TDRSS user community, this band would seem like a good candidate for future military expansion. This band is underutilized at this time.

25.25–27.00 GHz

Current Uses

International agreements limit the use of this band for links between satellites in space research and Earth exploration-satellite services, and for data from industrial and medical activities in space. The U.S. Navy is also a major user of this band for ship electronic testing.

Future Outlook

NASA is targeting this frequency band for future TDRSS expansion. With proper coordination, this band has potential for the Warfighter's Internet (UAV cross-link communications in particular).

30.0–31.0 GHz

Current Uses

This is one of several bands that has been allocated to military fixed and mobile satellite systems (although none operate within this band at this time).

Future Outlook

This band is already allocated for future military systems, and this band would seem like an easy and logical frequency spectra solution for the Warfighter's Internet

32.0–33.0 GHz

Current Uses

This band is used for some intersatellite and deep space uses.

Future Outlook

Potential future uses of this band include the Space Station for non-U.S. satellite connectivity. This band has strong potential for Warfighter's Internet, especially for UAV cross-link communications.

37.9.5–40.5 GHz

Current Uses

Television broadcast auxiliary stations use this band to connect remote and transmission sites. Government use of this band is limited to military satellites.

Future Outlook

This band also has potential for future military expansion for the Warfighter's Internet.

Project Frequency Allocations for the Warfighter's Internet

Frequency allocation is a major consideration in the overall design and architecture of the Warfighter's Internet. Appropriate frequency management must be provided to and from personal communication terminals as well as larger node mobile or stationary communication terminals. LPI/LPD characteristics are required on these links as well (20–30 dB processing gain). A frequency allocation for cross-link communications between multiple UAV nodes is required as well.

It would be most desirable to keep the frequency allocation, particularly for the air-to-ground and ground-to-air segments, as low as possible. This will keep the system design as simple as possible. However, the one major drawback of these lower frequencies lies more in the political arena, i.e., trying to obtain frequency clearance in already congested frequency bands. The following sub-sections provide one potential outlook on the frequency allocation choices for the Warfighter's Internet based mainly upon the previous discussions on frequency allocation.

Ground-to-Air Links (Personal Terminals)

This is the most critical link to keep as low in frequency as possible. The lowest frequency band that could potentially be used for this application is probably the 1429–1435 MHz band. Small, omnidirectional antennas on personal, handheld terminals should be able to be accommodated with data rates in the 10s of kbps. For example, multiple 50 kbps communications links with spread spectrum processing gain on the order of 20 dB for LPI/LPD could be accommodated for this link. Sectorized, higher gain antennas would be required for the receive, airborne portion of this communications link.

Ground-to-Air Links (Mobile and Stationary Terminals)

The lowest frequency band that could potentially be used for this application is the 4400–4500 MHz band with the 7.90–8.40 GHz band being a reasonable second choice. Choosing a frequency band as low as possible is not as critical for these types of terminals, as higher power amplifiers and directional or even tracking (in the mobile terminal case) antennas are feasible to support the higher data rate communication links. For example, either of these frequency bands could easily support multiple 500 kbps communications links with spread spectrum processing gain on the order of 20 dB (or more) for LPI/LPD. Once again, sectorized, higher gain antennas on board the UAV portion of this link would be required.

Air-to-Ground Links (Personal, Mobile, and Stationary Terminals)

The lowest frequency band that could potentially be used for this application is the 4400–4500 MHz band with the 7.25–7.75 GHz band being a reasonable second choice. As with the frequency selection for the ground-to-air links for the mobile and stationary terminal, choosing a frequency band as low as possible is not as critical for these links, as higher power amplifier and sectorized, higher gain antennas on board the UAVs can more than compensate for the higher expected propagation losses. For example, either of these frequency bands could easily support multiple 500 kbps communications links with spread spectrum processing gain on the order of 20 dB (or more) for LPI/LPD. Nevertheless, for personal terminals, a downlink lower in frequency (perhaps around 1435 MHz) is desirable for maximum foliage penetration and terrain variations.

UAV Cross-links

Either the 25.25–27.00 or the 32.0–33.0 GHz bands would make excellent choices for this type of communications link. Communication links on the order of 10s of Mbps could easily be supported at these frequencies with higher gain, tracking antennas. While LPI/LPD is not as critical for these links, 20 dB (or more) of processing gain could be accommodated as well.

APPENDIX H

ASSESSMENT OF CANDIDATE PROTOCOLS FOR WARFIGHTER'S INTERNET

Abstract

This appendix documents a brief study of data communication network protocols that could be adapted for use with Warfighter's Internet (WI). Our investigation leads us to recommend that the Space Communication Protocol Specification (or Standards)-Transport Protocol (SCPS-TP) be strongly considered for WI applications based on the following observations:

- The current plain vanilla version of the Transmission Control Protocol-Internet Protocol (TCP-IP), which was designed for operation over, and matched to the characteristics of, terrestrial, wireline, stationary, packet-switched (connectionless) networks, will not perform satisfactorily in a WI air-ground, wireless, mobile, connectionless environment without a suite of enhancements or extensions.

- To achieve maximum performance efficiency with minimum overhead, we would need to develop a completely customized protocol suite. This might be considered a viable option if our time and resources were unlimited, but this is not the case; nor is this approach consistent with the established goal of using commercial, off-the-shelf (COTS) technology where possible.

- We could select a set of protocol boosters (e.g., Snoop or Bellcore's "Protocol Boosters," both discussed below), but each such enhancement adds more overhead and increases the overall network latency. Furthermore, we would need to determine whether there is any degradation due to adverse interaction between these boosters (e.g., the negative impact of the TCP Vegas congestion control on the SCPS-TP reduced ACK frequency control, described below).

- SCPS-TP is a currently available ensemble of mutually compatible TCP enhancements with a relatively advanced level of maturity and heritage (i.e., development started in 1994 and is scheduled for completion by mid-1997) that can jump start the WI program: this will not necessarily be the only alternative at some later date, but it is certainly a protocol suite of opportunity at this point in time.

1. Introduction

The current Transmission (or Transport) Control Protocol and Internet Protocol (TCP-IP) standard was designed for, and is matched to, the characteristics of terrestrial, hard-wired, commercial data networks, including:

- Relatively noise-free (reliable) links, with loss of data primarily due to congestion
- Plentiful bandwidth
- Symmetric channel capacity
- Low propagation delays
- Static (stationary) backbone network (BBN) topography with continuous connectivity
- Computationally powerful end systems
- Point-to-point (unicast) and broadcast communication modes

By comparison, space/ground networks in general, and Warfighter's Internet (WI) in particular, feature:

- Noisy links with non-congestion-related data losses due to channel errors, link outages (e.g., resulting from handoff problems), jamming, and fading (e.g., multipath, shadowing)
- Limited bandwidth
- Asymmetric forward/reverse channel capacities
- Longer propagation delays
- Mobile intermediate switches and end users: dynamic network configurations with intermittent connectivity and low (e.g., 10%) contact duty cycle
- Computationally-limited end systems
- Multimedia network communications: data, video, and voice traffic
- Unicast, point-to-multipoint (multicast),[1] and broadcast communication modes

For example, it has been well documented (e.g., [1], [2]) that TCP-IP is optimized for congestion avoidance; that is, it responds to all data losses by reducing its transmission window size before retransmitting packets, invoking congestion control or avoidance mechanisms (e.g., slow start [3, p. 214-215]), and backing off its retransmission timer (Karn's algorithm [3, p. 212]). Unfortunately, when packets are lost for reasons other than congestion, as typified by wireless networks with high bit error rates (BERs), these measures result in suboptimal performance, i.e., an unnecessary degradation in end-to-end data throughput and very high interactive delays. TCP also does not perform well over links with long propagation delays and/or forward-reverse bandwidth asymmetries. Furthermore, it implements a *cumulative* positive acknowledgment (ACK) scheme as opposed to the arguably more efficient *selective* negative acknowledgment (SNACK) paradigm. Finally, under degraded channel conditions, TCP tends to generate an excessive amount of overhead traffic, with a corresponding decrease in throughput efficiency.

[1] Multicasting is the ability of one end user to send replicas of the packets comprising a given message to a group of other end users. It is a function that has led to the development of address resolution protocols (ARPs).

The User Datagram Protocol (UDP) is sometimes substituted as an unreliable alternative to TCP. It has less overhead and is therefore faster than TCP because it has no provisions for acknowledgments. UDP functions best in an environment in which short messages need to be transmitted between end users on a best efforts basis. UDP also has some segmentation and reassembly issues that need to be addressed. UDP in conjunction with IP currently offers an unreliable multicast capability; however, it is believed that WI will require reliable multicast service.

If we accept the fundamental premise that plain vanilla TCP-IP has some serious deficiencies for WI applications, what is the best approach to alleviating these in the short term? Since WI is an autonomous network that does not have to necessarily conform to pre-existing commercial standards, the optimum solution from a purely academic perspective is to carefully measure the relevant system parameters (e.g., propagation delays, channel error rates, mobility, etc.) and develop a completely customized "Warfighter's Internet Protocol" uniquely matched to these (possibly time-varying and stochastic) characteristics. In principle, this would result in the leanest, integrated system controls with no unnecessary overhead and minimum latency. However, this is not a practical answer given the real-world limitations imposed on us by time and resource considerations. Furthermore, we run the risk of producing a protocol architecture that is too application-specific, denying us the flexibility to make unanticipated, fundamental changes to our communication system design later on while maintaining near optimum performance efficiency. Also, a uniquely customized set of protocols may not be very robust to system parameter variations, resulting in significant mismatch degradation when some of these network attributes drift from their nominal specifications. Finally, a roll-your-own strategy is inconsistent with the established goal of using commercial, off-the-shelf (COTS) technology wherever appropriate.

So, with these considerations, we are forced to adapt some existing alternatives. The primary available options are wireless asynchronous transfer mode (ATM), possibly with encapsulated IP, protocol enhancements or "boosters," and the Space Communication Protocol Specification (SCPS) suite of integrated protocols; the merits (and demerits) of each technology for WI are reviewed below.

2. Wireless ATM

ATM technology (e.g., [4]–[6]), which is often interpreted to be synonymous with the Broadband Integrated Services Digital Network (B-ISDN), is fundamentally a connection-oriented, virtual circuit (VC), scaleable fabric designed for high-speed (i.e., many Gigabits/sec), multimedia, unicast communications with relatively low latency. The Army already has a significant commitment to the use of ATM technology for legacy radio networks such as their Mobile Subscriber Equipment (MSE), and for future systems such as Radio Access Point (RAP) communications concentrator nodes currently being developed.

Each ATM cell consists of a short, fixed-length packet containing a 5-byte header followed by 48 bytes of data. The ATM header contains a virtual circuit identifier (VCI) and a virtual path identifier (VPI), where virtual paths are higher-level aggregations of VCs. The VPI/VCI labels only have local significance associated with a particular link, and each ATM switch can introduce its own input/output VPI/VCI mapping. Unlike packet-switched (connectionless) systems, there is no provision for misordered received cells; data on each VC is assumed to be received in the original transmitted sequence. This simplified structure is a major reason for the intrinsically high-speed/low-delay capability of ATM switches, which do not have to process variable-length, generally longer packets (e.g., several kilobytes) with headers containing end-to-end routing information that are characteristic of the relatively more complex, connectionless, shared-resource networks. However, this same restriction to short, fixed-length 53-byte cells with 5-byte headers that allows ATM to enjoy the benefits of high transmission rates with minimal delays is also a potential source of performance degradation; the header-to-data ratio represents an underlying 10% overhead, and the packet length may be too short for optimum data throughput efficiency at lower bit error rates (BERs) [7].

Unlike connectionless networks, ATM requires a separate signalling channel to establish VCs prior to each communication session; this can represent a significant additional overhead for short message transfers over and above the fixed 10% data transfer overhead. ATM operates in a layered architecture similar to TCP-IP; i.e., the ATM protocol reference model has some resemblance to the 7-layered International Standards Organization (ISO) Open System Interconnection (OSI) reference model, with communication from higher layers occurring through three lower layers—the ATM adaptation layer (AAL), the ATM layer, and the physical layer (in descending order).

Although ATM is primarily attuned to connection-oriented traffic, including Class A (AAL Type 1) constant bit rate uncompressed voice or video, Class B (AAL Type 2) variable bit rate compressed voice and video, and Class C (AAL Type 3/4 and AAL Type 5) variable bit rate data (e.g., X.25), it will also work with Class D (AAL 3/4) connectionless packet data such as local area network (LAN, e.g., Ethernet) traffic. For example, ATM systems can handle IP traffic through the use of tunneling (encapsulation) techniques whereby an IP packet is encapsulated within an ATM cell; however, if the IP packet is longer than an ATM cell, it will require segmentation and reassembly (SAR) of the IP packets which will increase the end-to-end latency.

Because ATM is intrinsically connection-oriented and point-to-point (unicast), multicasting and broadcasting are not intuitive functions in an ATM network. To achieve a multicast capability within an ATM architecture, mechanisms need to be created to establish some form of multipoint connection between all nodes in a multicast group, and to replicate cells at ATM switch interfaces. The ATM Forum is currently addressing this first deficiency.

The replication issue is related to the type of ATM switch used. Most conventional, matrix-type ATM switch implementations are incapable of performing both the cell routing and replication functions. This implies that a separate "copying fabric" needs to be prepended to the switch matrix. However, the recently developed backplane-based ATM switches [4, pp. 54–57], by virtue of their bus architecture, can perform multicasting more readily because in a bus, as on a LAN, each node sees all traffic and can simply copy multicast cells.

One of the principal deficiencies that mitigates against the adoption of wireless ATM for WI applications is its reliance on high quality wireline (e.g., fiber optic) links with nominal BERs no greater than 10^{-9} [5, p. 4]. For links with higher BERs, reliable wireless ATM communications would entail frequent retransmissions or a powerful, low-rate forward error correction (FEC) code, which could degrade end-to-end performance, especially the overall latency.

Because it was designed primarily for operation over low loss, fiber optic media, ATM has minimal built-in capability for dealing with bit errors. For example, there are no automatic repeat request (ARQ) provisions at the ATM physical layer (although ARQ could be implemented at the AAL), and only the header (i.e., not the data) is protected by a forward error correction (FEC) code.

Without the intrinsic protection against undetected entry by non-intrusive means afforded by fiber optic links or other cable-based connections, wireless ATM is inherently vulnerable to interception and jamming (including spoofing). There are no imbedded security measures which might inhibit the extraction of intercepted information bits, particularly in light of its open architecture tied to the given AAL mode. Once the desired ATM traffic has been intercepted by an adversary, the AAL type may be readily recognized by determining the order of the cyclic redundancy check (CRC) within the sequence number protection (SNP) field of each header [5, pp. 39–59]. For example, the SNP for AAL Type 1 service has an even parity 3-bit CRC, AAL Type 3/4 has a 10-bit CRC, and AAL Type 5 uses a 32-bit CRC (the CRC for AAL Type 2 has not yet been specified). It should be noted, however, that because the VPI/VCI numbers have only local link significance, unless the intercepted traffic is from the first or last network hop, the problem of identifying the end users still remains. Also, although not encrypted, if the data bits are scrambled as is recommended in the ATM physical layer specification to alleviate the potential for false synchronization, this will simultaneously provide some measure of information security.

A detailed discussion of the ATM physical layer vulnerabilities intrinsic to the cell synchronization function, header error control (HEC) algorithm, and the cell scrambling operations is given in [5, pp. 47–50], followed by a revealing assessment of corresponding AAL weaknesses to both random bit errors and specific error patterns in [5, pp. 50–58].

Finally, if the network topology is dynamic and time-varying as is expected for WI, there may be a need to reroute VCs within a given message session, again increasing latency.

For all of these reasons, although ATM should be supported by WI through gateways to accommodate legacy radios, we would not recommend the adoption of wireless ATM as the fundamental WI protocol of choice, even with provision for encapsulating IP traffic over ATM networks.

3. Protocol Boosters

Another approach to alleviating some of the deficiencies in TCP-IP for wireless data communications in general and WI applications in particular is the use of protocol "boosters" (i.e., enhancements or extensions to TCP-IP). TCP enhancement techniques, in particular, are generally designed to remedy the behavior of the core transport protocol to non-congestion related losses characteristic of wireless networks. They involve a variety of mechanisms, including local retransmissions such as Snoop [8], split-connection schemes such as Indirect TCP (I-TCP) [9], and FEC to improve end-to-end throughput performance. An excellent overview of the weaknesses of TCP and a comparison of several TCP enhancements is presented in [10].

The advantage of protocol boosters is that they allow the retention of the basic underlying protocol by the addition of a relatively minor local modification ("band-aid") that addresses a specific TCP-IP deficiency. Consequently, they can, in theory, be developed much more quickly than a new global (end-to-end) protocol. Furthermore, they can, again in theory, be tailored to a particular environment, although this could actually result in a mismatch degradation if the network attributes are either difficult to measure or vary too rapidly to adapt to, and the performance with the protocol extension is sensitive (i.e., non-robust) to these system variations.

It should be noted that each such protocol enhancement introduces additional overhead and network latency. Each booster needs to "touch" the protocol data units (PDUs) to a lesser or greater extent; this handling of the PDUs contributes to the overall latency, and the less the booster needs to interact with the data, the more efficient the implementation. For example, an FEC booster might add *n-k* parity packets to every *k* data packets, which immediately increases the overhead by a factor of *n/k*, the inverse of the code rate. Also, in general, adding a "band-aid" to an existing protocol is a suboptimum solution in comparison with an integrated protocol suite. While a particular booster may locally solve a point problem, it may interfere with other aspects of the core protocol. As the number of core protocol deficiencies alleviated by boosters increases, so does the complexity of configuring and maintaining the resulting system. This increased complexity can lead to unexpected performance degradations that are not immediately evident in simpler configurations:[2] there is always the possibility of adverse interactions between two or more boosters (see discussion below).

As an example of the benefits that can be derived from a booster, consider the Snoop protocol referred to earlier. This is a link-layer mechanism that incorporates and exploits its knowledge of the higher-layer transport protocol, TCP. As noted in [10], "the main advantage of employing a link-layer protocol for loss recovery [in a wireless network] is that it fits naturally into the [ISO-OSI] layered structure of network protocols. The link-layer protocol operates independently of higher-layer protocols (which makes it applicable to a wide range of scenarios), and consequently, does not maintain any per-connection state. The main concern about link-layer protocols is the possibility of adverse effect on certain transport-layer protocols such as TCP."

The Snoop protocol employs a so-called Snoop Agent collocated with the router in the source node base station to (1) monitor TCP traffic in both directions, and to (2) maintain a queue of outgoing TCP packets for the purpose of local retransmission of unacknowledged packets. The Snoop Agent detects a packet loss by the arrival of duplicate acknowledgments from the destination node or by a local time-out prior to the receipt of an acknowledgment. In either event, it (1) suppresses the duplicate acknowledgments, and (2) retransmits the lost packet from its local cache. This ensures that the source does not attempt to invoke TCP congestion control or

[2] Bob Durst and Eric Travis likened the problem of combining boosters with "the mixing of different medications. Each medication is a focused approach to treating a specific ailment. As the number of medications ingested increases, so does the danger of harmful interactions between the different remedies..."

unnecessarily enter the fast-retransmission mode when the losses are related to corruption of the wireless links.

Snoop has been demonstrated to be even more effective when combined with some form of selective acknowledgment (SACK)[3] such as the Smart scheme [12] as opposed to the cumulative acknowledgments (ACKs) used by conventional TCP (refer to [10], Figure 7 for a graphical description of Snoop and Smart, and Figure 11 for the relative throughput performance benefits).

Bellcore is also in the process of developing and testing its own proprietary suite of boosters that extends the capabilities of TCP-IP (but not SCPS) [13–15]. The Bellcore "Protocol Boosters" will reputedly allow end users to continue to experience the benefits of TCP-IP while enhancing only the wireless portion of the network. It is designed for application to personal communications systems (PCS), cellular telephony, low-Earth orbit (LEO), and geostationary (GEO) orbiting satellite-to-ground communications, excluding deep space applications. Some of the technology has already been tested over satellite and terrestrial wireless links.

The Bellcore Protocol Boosters suite will also include a capability for reliable multicast, a feature of particular interest to WI. In fact, a protocol booster (not necessarily Bellcore's version) may be the most expedient way to deal with multicast for the mobile WI environment. For example, we could implement a "selective forward" function which provides for the local replication and forwarding of certain multicast messages from a particular radio access concentration node to all connected members of the multicast group.

As a precautionary note, it should be mentioned that several knowledgeable sources in the field of communication networks have expressed some skepticism over the claims made by Bellcore regarding the alleged capabilities of their Protocol Boosters. It is not immediately clear whether these doubts represent objective and informed assessments or are motivated by other, less scientifically legitimate reasons.

4. Potential for Adverse Interactions Between Boosters

Extreme caution must be taken to ensure that inter-booster interactions are well understood. For example, in the vernacular of [13, Sections 2.4.1 and 2.5.2], it is not clear what would happen to UDP traffic if one used a two-element FEC booster at the edges of a wireless network, and added a one-element UDP error detection booster somewhere in the middle of the network.

UDP headers contain a 16-bit checksum field that may or may not be used. If unused, the value is 0x0000. The 1-element UDP booster simply computes and inserts a 16-bit checksum into the UDP header if it is not already used. The first element of the 2-element FEC booster pair computes FEC bits and ships them (independent of the data packets?) to the second element, which uses the FEC bits to help decode the data packet.

[3] According to [10]. SACKs were added as an option to TCP by Request for Comments (RFC) 1072 [11], but later disagreements over their use prevented the subsequent universal adoption of this specification.

To further illustrate this point, consider a situation like that in Figure 1. If the fixed terminal is communicating with the portable terminal using UDP without using the checksum option,[4] and if some node in the middle of the wireless network decides to invoke the 1-element UDP booster, which is not unreasonable, then the mobile user will decode every packet in error, since the FEC bits are computed based on a UDP checksum of 0x0000 which is *not* the checksum of the received packets.

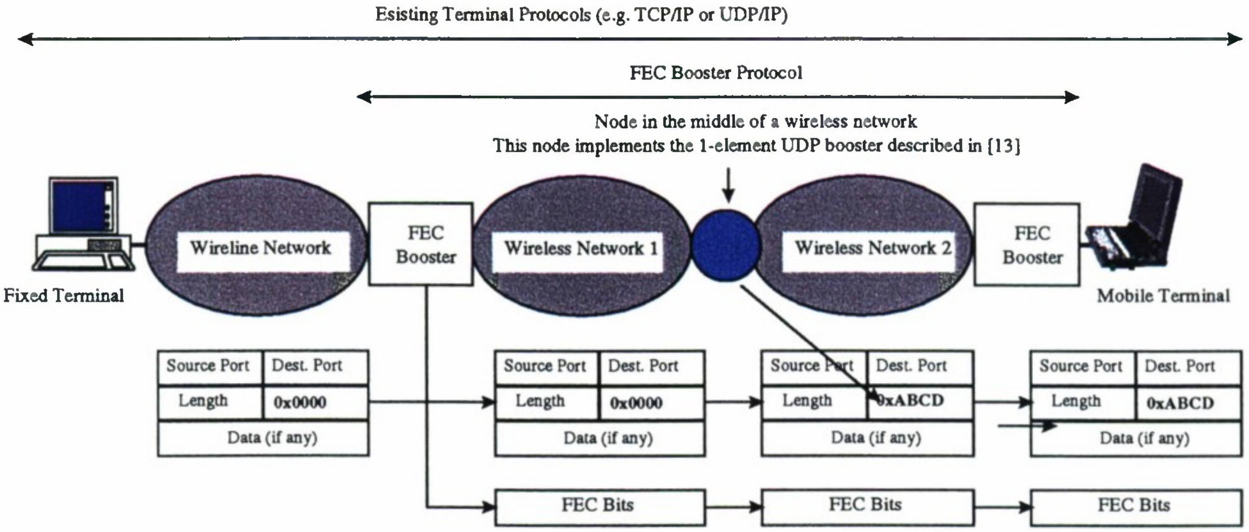


Figure 1. Example use of UDP booster protocol.

Granted, the description of the 1-element UDP booster in [13] describes its use for packets "leaving a reliable LAN," not for use in the middle of a wireless network. Then the problem posed in Figure 1 derives from the order in which the boosters are applied at the gateway to the wireless network. If the UDP booster is applied first, everything works satisfactorily; if the FEC booster is applied first, the problem occurs.

This is admittedly a contrived example, but one that uses the existing booster descriptions to illustrate that simply slapping boosters on the problem cannot be done with abandon; some care must be taken. One might find similar problems using the 2-element jitter control booster [13, Section 2.5.1] with TCP Vegas, for example, as TCP Vegas depends on accurate measurements of the round-trip times (RTTs) to clock packets into the network. The degradation would be less severe, since packets would probably not be lost, but some of Vegas' improved throughput would probably be sacrificed.

[4] Some implementations of the Network File System (NFS) use UDP.

5. SCPS-TP

SCPS is a collection of four layered protocols that can function either as a single, integrated (ISO/OSI Layer 3-7) stack, or be individually selected to match a particular application of interest [2, 16–23]. SCPS is intended to operate over existing space, telemetry, telecommand and communication (TT&C) channels, such as the standard Consultative Committee for Space Data Systems (CCSDS) link layer, or a space-ground link system (SGLS), and are specified in a set of CCSDS recommendations.

SCPS has been designed to accommodate an environment in which one end user is on the ground and the other end user is hosted in a spaceborne computer with relatively constrained processing and memory capabilities and where communication bandwidth is at a premium; the code implementations are therefore deliberately "skinny" and operate with high data transfer efficiency. The same attributes that serve SCPS so well for ground-space applications also make it suitable for other "high stress" environments,[5] including multi-hop, wireless, mobile, tactical networks such as WI. Different combinations of these stressed environment characteristics are exhibited in applications ranging from terrestrial cellular and land-mobile radio networks to geostationary (GEO) and deep-space satellites. Because of this wide scope of requirements, and the need to keep the resulting protocols compact and processing friendly, all of the SCPS protocols are inherently flexible, modular, and scaleable. Core functionality is deliberately kept small, but augmented with a menu of many optional capabilities, each of which can be individually invoked if required by the operating environment; a user need only support as much of the SCPS protocol (and overhead) as required. These optional SCPS features are only enabled if their use is negotiated by both end users. For example, SCPS-TP (Transport Protocol) reverts to conventional TCP if both end-points of a transport connection do not flag the SCPS option in their initial protocol handshaking.

The SCPS protocol suite includes:

(1) SCPS-FP (File Protocol) – This is a "tuned-up" version of the conventional Internet File Transfer Protocol (FTP) operating at ISO/OSI Layer 7. It is optimized to operate more efficiently than FTP in a space environment (e.g., to handle functions such as uploading spacecraft commands and software, and downloading telemetry data) and to provide certain new services that are required by space missions (such as the ability to manipulate individual records without reloading the whole file). It is interoperable (i.e., backward compatible) with commercial FTP at the cost of some of the enhanced capabilities. The "Noninteractive File Transfer Protocol" (SCPS-NiFTP) is a variant of the SCPS-FP that is also under development; this new protocol is not FTP-interoperable, but it can tolerate long propagation delays without the need for time-consuming handshaking between space and ground nodes.

[5] The SCPS design team characterizes stressed environments as having features similar to the bulleted list of WI characteristics given in the Introduction, which will not be repeated here. Some of the attributes of SCPS presented in this section are extracted from an e-mail message sent to me last January by Bob Durst and Eric Travis, who are in charge of the SCPS-TP development effort.

(2) SCPS-TP (Transport Protocol) – This is an enhanced version of TCP/UDP operating at Layer 4 that is reviewed in more detail below.

(3) SCPS-SP (Security Protocol) – This optional data protection mechanism provides selectable levels of end-to-end security (e.g., message authentication, access control, integrity, and encryption) at Layer 3.5.

(4) SCPS-NP (Network Protocol) – This provides functionality similar to IP at Layer 3. However, it has been customized to support unique routing configurations (e.g., flood routing through constellations of spacecraft) and has been "shrunk down" for space networking applications (characterized by fewer nodes relative to global networks) so as to minimize overall communications overhead. It can operate in a connection-oriented (circuit-switched) mode similar to the current CCSDS "Path" service, or it can support connectionless (packet-switched) IP-like routing. It is a fully scaleable protocol, whose features (and therefore overhead) are selectable to match the application (e.g., address sizes that can range from no address all the way to IP Version 6).

The SCPS protocols and their accompanying software are a joint NASA–Department of Defense (DoD) development which began in 1994 and is scheduled for completion this summer. At that time, SCPS-TP (and possibly SCPS-SP) will be ported to a JPL testbed where it (they) will be available for simulated performance trials. SCPS therefore has a level of maturity and heritage that currently is ahead of the development cycle of any other protocol suite of opportunity, including Bellcore's Protocol Boosters.

The SCPS project is jointly managed by NASA (Adrian Hooke, JPL) and by the USAF Space and Missile Systems Center (SMSC). From the outset, the project followed a policy of outsourcing: the SCPS-FP is being developed under contract by Science Applications International Corporation (SAIC); the SCPS-SP by Howard Weiss of SPARTA; and the SCPS-TP and SCPS-NP by Bob Durst at MITRE (which is also responsible for overall stack integration and testing) and Eric Travis, a former MITRE employee now affiliated with Gemini Industries.

Of the entire suite of four SCPS protocols, SCPS-TP is the best candidate for adoption by WI, with SCPS-SP and/or SCPS-NP also possibilities. SCPS-TP includes a mixture of several extensions/options to TCP, the advantages of which are summarized below and described in detail in [2]:

(1) TCP Timestamps and Window Scaling options to handle long delays and high volumes of in-transit data

(2) Reduced Acknowledgment Frequency Control to accommodate channel asymmetries

(3) A complete specification of previously defined but as yet undeveloped TCP optional capabilities (e.g., SNACK, Header Compression)

(4) Some new capabilities, such as a "Best Efforts" service which continues to deliver data (under a so-called "Persist Mode") even if the acknowledgment channel becomes temporarily unreliable.

The choice of SCPS-TP options affect its interoperability with conventional TCP. The use of these options is negotiated on the establishment of a connection via an "SCPS-TP capable" flag. If the other peer TCP end user does not return that notification on the SYN segment, SCPS-TP reverts to standard TCP operation.

The loss mechanisms of SCPS-TP have extended capabilities relative to TCP. Unlike TCP, the SCPS-TP default assumption can be set by the user or the application. This assumption can be overridden by an explicit signal identifying the particular source of loss; for example, a corruption-experienced Internet Control Message Protocol (ICMP) notification can be sent to the destination node, which then triggers a corruption-experienced option in its ACK. There are also link-outage ICMPs and the previously mentioned Persist Mode.

To cope with asymmetric channels (i.e., different bandwidths or data rates for the forward and reverse links), SCPS-TP can remove the conventional TCP requirement to generate one ACK at least every other segment. There is also a reduced acknowledgment frequency control, which is currently running in an open loop mode for which it needs substantial tuning because of its interaction with SCPS' TCP Vegas congestion control; there are plans to remedy this sensitivity by moving to a closed-loop tuning mechanism and developing adaptive ACK strategies.

For links with limited capacity (i.e., bandwidth), SCPS-TP employs end-to-end, variable-length TCP (not IP) header compression, which is negotiated by the source and destination end users via an option on the uncompressed SYN segments. SCPS-TP does not use the delta-encoding scheme specified in TCP-IP's RFC 1144, which is sensitive to corruption, loss, and misordered packets. Instead, the SCPS-TP header compression function succinctly summarizes static information and omits any other irrelevant data.

As noted earlier, SCPS-TP uses the SNACK option as opposed to standard TCP cumulative ACKs. SNACK borrows concepts from selective positive acknowledgments (SACKs), which are less bandwidth-efficient, and negative acknowledgments (NAKs), which can only identify a single hole in sequence space. The variable-length option carried on the ACK segment can identify multiple holes in a receiver's out-of-sequence queue.

SCPS-TP also employs two TCP options from RFC 1323, including:

(1) Timestamps, which provide more accurate RTT estimates

(2) Window Scaling, which enables window sizes larger than 64 kbytes (with corruption as opposed to congestion, a window greater than the bandwidth-delay product allows the source to continue transmitting new data during any loss recovery phases).

6. Performance Comparison Tests Between TCP and SCPS-TP

As described in [2, 24–26], a number of tests were conducted using an Ethernet testbed and a satellite channel simulator ("Spanner") in the laboratory, and using over an actual satellite link, to compare the performance of a prototype version of SCPS-TP with regular TCP in various controlled environments. The TCP implementation was incorporated into a SunOS 4.1.3 Unix kernel; this version does not include the RFC 1323 extensions above, and has a maximum window size of 51 kbytes. To ensure that the TCP throughput results were not biased by this reduced window size, all tests were performed on a network with a bandwidth-delay product below the SunOS limit. The TCP Timestamps and SCPS-TP Header Compression options were both disabled so that both protocols had comparable header overheads (except for the SNACK option).

The Ethernet laboratory testbed and the Spanner satellite channel simulator were used to compare the behavior of TCP and SCPS-TP to link bandwidth asymmetries [2]. In particular, the tests were performed with a 50 ms one-way propagation delay (100 ms RTT), and a 1.5 Mb/s (T1) data rate in the forward direction. The ACK return channel data rate was varied from 1.5 Mb/s down to 1.5 kb/s to create forward-return link asymmetry ratios ranging from 1:1 to as low as 1000:1. Note that in all cases, the bandwidth-delay product of the network did not exceed 1.5 Mb/s x 100 ms = 150 kb = 18.75 kbytes, which is well below the SunOS TCP limit of 51 kbytes. Also, the transmit and receive buffer sizes were each set to 300 kbytes; the SCPS-TP Congestion Control, Window Scaling, and SNACK (with no bit-vector) options were enabled; the TCP Timestamps and SCPS-TP Header Compression options were disabled.

The results of the asymmetry simulation tests are shown in Figure 2 below, which has been excerpted from [2] with the permission of the authors. We see that on a symmetric (1:1) channel, TCP and SCPS-TP both produce nearly the same average data throughput. However, as the ACK (return) channel capacity is reduced in comparison with that of the data (forward) channel, TCP's throughput suffers an apparently exponential decay while SCPS-TP degrades almost linearly and certainly much more slowly.

The local dip in the SCPS-TP throughput at an asymmetry ratio of 400:1 illustrates an apparent tuning mismatch sensitivity of the protocol due to an adverse interaction between the TCP Vegas congestion control mechanism that was adopted by SCPS-TP and the acknowledgment strategy.[6] If the reduced acknowledgment frequency is not tuned to match the ACK channel bandwidth, the TCP Vegas congestion control algorithm assumes that the ACKs are getting backed up due to congestion and reduces the data throughput unnecessarily. This was, in fact, what happened in the 400:1 test case resulting in the observed suboptimal throughput reading. By comparison, TCP's congestion control algorithms are much more robust, and its throughput degrades smoothly as the asymmetry ratio increases. The developers of SCPS-TP plan to enhance their implementation of TCP Vegas congestion control by providing a closed-loop tuning mechanism and introducing adaptive ACK strategies to address the combination of high error rates and limited ACK channel bandwidth.

[6] This is also exemplary of the kinds of unexpected interplay and corresponding degradation that can occur between two protocol boosters.

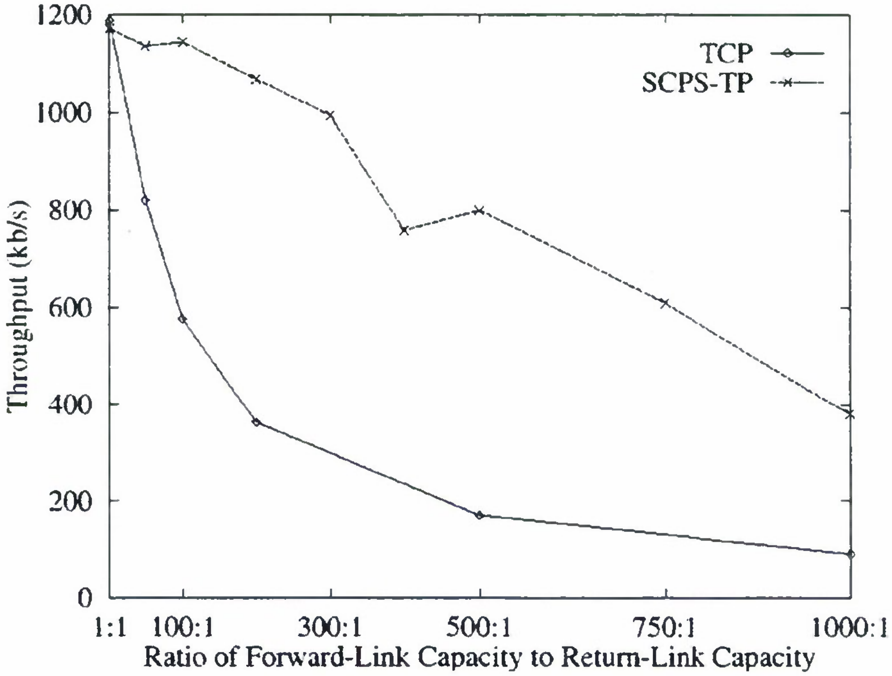


Figure 2. Asymmetric channel: SCPS-TP vs. TCP.

The next set of tests using the Spanner satellite channel simulation application and the Ethernet laboratory testbed looked at the impact of link data errors or corruption on the relative average data throughput efficiency for comparable implementations of TCP and SCPS-TP [2]. This time, the simulated environment included symmetric 1.5 Mb/s forward and return channels, 50 ms and 100 ms one-way propagation delays (bandwidth-delay products of 18.75 kbytes and 37.5 kbytes, which is again less than the SunOS's TCP limit of 51 kbytes), and transmit and receive buffers each set to 300 kbytes. The Window Scaling and SNACK options were enabled, while the Congestion Control, Header Compression, and Timestamps options were turned off.

Results were obtained for BERs ranging from 10^{-8} to 10^{-5} (Figure 3 below, excerpted from [2]). It is evident that the average data throughput for TCP degrades severely as the BER increases for both simulated delays, whereas the corresponding performance for SCPS-TP remains almost unaffected by the changing link corruption until the BER falls below 10^{-6}. The reasons for this behavior on the part of TCP have been discussed before in this paper, and are well documented in [1, 2]; the default congestion control bias of TCP reduces its window size unnecessarily when corruption is experienced, and in the absence of a selective acknowledgment option, TCP learns about the packet losses relatively slowly, i.e., at most one packet loss per RT.

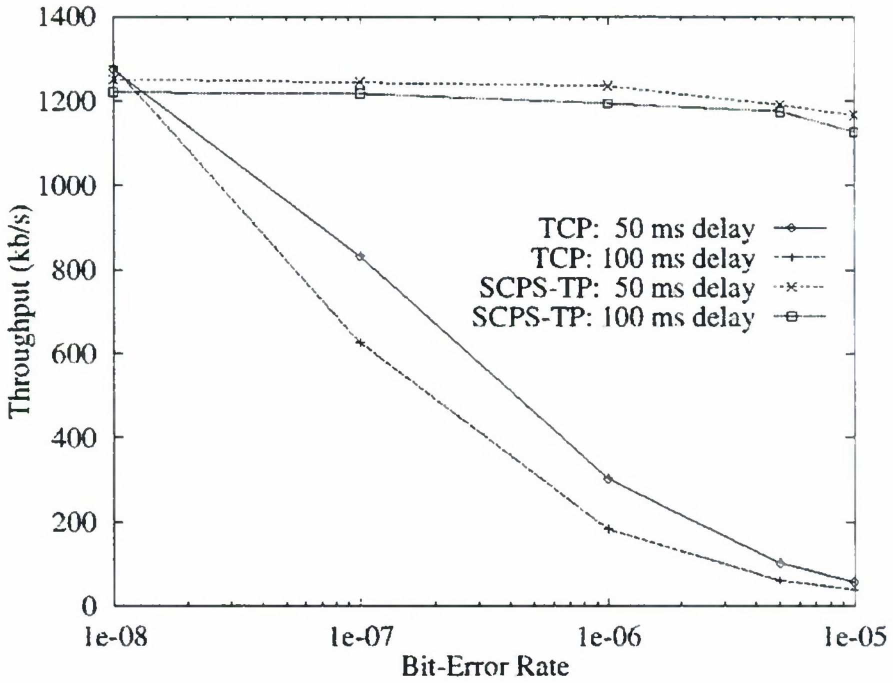


Figure 3. Corrupted link: SCPS-TP vs. TCP.

For the bent-pipe experiments, SCPS-TP was implemented over an actual U.S. DoD satellite link, and the results were then compared with laboratory data [2, 24]. Test conditions included a 256 kb/s forward direction data rate and a 32 kb/s ACK return channel (8:1 asymmetry ratio); 500 ms RT (GEO) propagation delay (bandwidth-delay product equal to 128 kb or 16 kbytes); and transmit and receive buffers, each set to approximately 70 kbytes. Also, the SCPS-TP loss assumption default condition was set to corruption; the Window Scaling, Timestamps, Header Compression, and SNACK options were turned on; Congestion Control was turned off.

The average data throughput is plotted as a function of the BER for 1000-byte and 125-byte segments in Figure 4 (excerpted from [2]). We see that there is excellent correlation between the actual bent-pipe satellite measurements and the laboratory simulations. Also, the throughput performance is virtually unaffected by BERs lower than 10^{-6} for the longer packet lengths and 10^{-5} for the shorter ones. Finally, as a demonstration of the impact of the packet length on throughput performance (refer also to the Appendix in this report on "Derivation of Optimum Packet Length for ARQ Wireless Data Communications"), notice that the 1000-byte segments outperform the 125-byte segments until the BER exceeds 10^{-5}, at which point the shorter packet lengths are preferred because they are less susceptible to packet errors in a corruption-dominated environment.

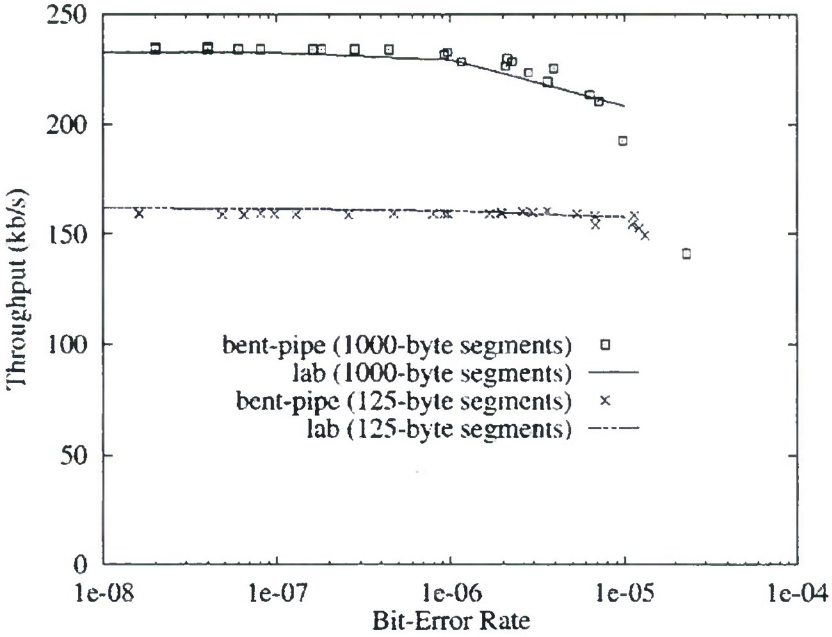


Figure 4. Bent-pipe test: throughput vs. BER.

The last test sequence involved the implementation of SCPS-TP on board a United Kingdom STRV satellite [2, 25–26] and in the laboratory simulation testbed. Relevant test parameters included a forward direct data rate of 1000 b/s and an ACK return link of 125 b/s (8:1 asymmetry ratio); an 8 s RT propagation delay (larger than desired due to real-world clocking and ground component delays, but still achieving a bandwidth-delay product of 1 kbyte); a 19.712 kbyte on-board transmit buffer, a 69.88 kbyte ground receive buffer, and a 90-byte maximum transfer unit (MTU) limitation imposed by the STRV platform. Also, the Window Scaling option was enabled, while the Congestion Control and Timestamps options were disabled; tests were conducted with the SNACK and Header Compression options both on and both off.

The results are illustrated below in Figure 5 (reproduced from [2]). As expected, the advantages of the SNACK and Header Compression options are clearly evident. And, once again, the laboratory simulations closely track the actual satellite observations.

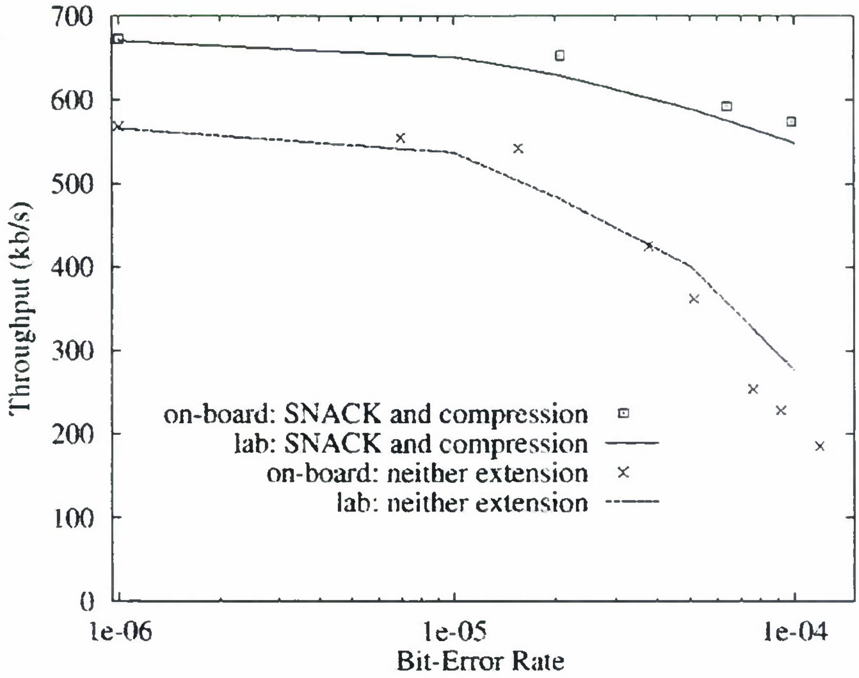


Figure 5. On-board test: throughput vs. BER.

7. Conclusions

Of the data communication network protocols that are currently available for use with WI, the SCPS protocol suite in general, and SCPS-TP in particular, come closest to satisfying our ensemble of requirements. A fully customized protocol stack is impractical due to cost and time constraints; wireless ATM, with or without encapsulated IP packets, will not perform adequately at moderate to high BERs because of its design limitations (e.g., lack of provision for ARQ and FEC, and fixed, short packet lengths); protocol boosters are essentially a nonintegrated approach that only provides a temporary band-aid solution with the potential for adverse inter-booster interactions.

References

1. Hari Balakrishnan, Venkata N. Padmanabhan, Srinivasan Seshan, and Randy H. Katz, "A Comparison of Mechanisms for Improving TCP Performance over Wireless Links," *Proceedings of ACM SIGCOMM '96*, Stanford, California, August 1996.

2. Robert C. Durst, Gregory J. Miller, and Eric J. Travis, "TCP Extensions for Space Communications," *Proceedings of the 2nd ACM Conference on Mobile Computing and Networking (Mobicom '96)*, November 1996.

3. Douglas E. Comer, *Internetworking with TCP/IP, Volume I: Principles, Protocols and Architecture, Third Edition*, Prentice Hall, Upper Saddle River, New Jersey, 1995.

4. Anthony Alles, "ATM in Private Networking," Tutorial, Hughes LAN Systems, Mountain View, California, 1993.

5. Martin J. Agan, Thomas C. Jedrey, and Richard P. Romer, "Asynchronous Transfer Mode (ATM): Overview, Applications, and Vulnerabilities," Technical report, Jet Propulsion Laboratory, Pasadena, California, March 31, 1995.

6. T. Russell Hsing, Donald C. Cox, Li Fung Chang, and Thierry Van Landegem, eds., "Wireless ATM," *IEEE Journal on Selected Areas in Communications*, Vol. 15, No. 1, January 1997.

7. Barry K. Levitt, "Derivation of Optimum Packet Length for ARQ Wireless Data Communications," Technical memorandum, Jet Propulsion Laboratory, Pasadena, California, February 18, 1997.

8. Hari Balakrishnan, Srinivasan Seshan, and Randy H. Katz, "Improving Reliable Transport and Handoff Performance in Cellular Wireless Networks," *ACM Wireless Networks*, December 1995.

9. A. Bakre and B. R. Badrinathe, "I-TCP: Indirect TCP for Mobile Hosts," *Proceedings of 15th International Conference on Distributed Computing Systems (ICDCS)*, May 1995.

10. Hari Balakrishnan, Venkata N. Padmanabhan, Srinivasan Seshan, and Randy H. Katz, "A Comparison of Mechanisms for Improving TCP Performance over Wireless Links," *Proceedings of ACM SIGCOMM '96*, Stanford, California, August 1996.

11. V. Jacobson and R. T. Braden, "TCP Extensions for Long Delay Paths," *RFC 1072*, October 1988.

12. S. Keshav and S. Morgan, "Smart Retransmission: Performance with Overload and Random Losses," *http://www.cs.att.com/netlib/att/cs/home/keshav/papers/smart.ps.Z*, Preprint, 1996.

13. D. C. Feldmeier, A. J. McAuley, and J. M. Smith, "Protocol Boosters," Bellcore and University of Pennsylvania, submitted to *IEEE Journal on Selected Areas in Communications (JSAC)*, Draft publication received January 1997.

14. A. J. McAuley and D. C. Feldmeier, "A Powerful Forward Erasure Correction Code Suited to High-Speed Software Implementation," Bellcore, Draft technical memorandum, November 14, 1996.

15. A. J. McAuley, D. S. Pinck, T. Kanai, M. Kramer, G. Ramirez, H. Tohme, and L. Tong, "Experimental Results from Internetworking Data Applications over Various Wireless

Networks Using a Single Flexible Error Control Protocol," in *International Journal of Satellite Communications*.

16. Robert C. Durst, "SCPS Network Protocol Specification Workshop," MITRE, Overheads, July 13, 1994.

17. Robert C. Durst, "Scaleable SCPS-NP," MITRE, Overheads, September 21, 1994.

18. Capt. Dave Syndergaard, "The Space Communications Protocol Standards (SCPS) Project," HQ NORAD/USSPACECOM J4, Overheads from SATCOM Interoperability and Standards Committee (SISC) Meeting, January 1996.

19. "Space Communications Protocol Specifications (SCPS) – Rationale, Requirements and Application Notes," Revised Draft "Green Book" Report Concerning Space Data System Standards, Consultative Committee for Space Data Systems (CCSDS) 710.0-G-0.2, August 1996.

20. "Space Communications Protocol Specifications (SCPS) – File Protocol (SCPS-FP)," Draft "White Book" Recommendation for Space Data System Standards, Consultative Committee for Space Data Systems (CCSDS) 717.0-W-1, January 1996.

21. "Space Communications Protocol Specifications (SCPS) – Transport Protocol (SCPS-TP)," Revised Draft "Red Book" Recommendation for Space Data System Standards, Consultative Committee for Space Data Systems (CCSDS) 714.0-R-1, May 1996.

22. "Space Communications Protocol Specifications (SCPS) – Security Protocol (SCPS-SP)," Revised Draft "Red Book" Recommendation for Space Data System Standards, Consultative Committee for Space Data Systems (CCSDS) 713.5-R-1, May 1996.

23. "Space Communications Protocol Specifications (SCPS) – Network Protocol (SCPS-NP)," Draft "White Book" Recommendation for Space Data System Standards, Consultative Committee for Space Data Systems (CCSDS) 713.0-W-1, January 1996.

24. "Space Communications Protocol Standards (SCPS) Bent-Pipe Experiment Report," Technical Planning Report SCPS-D71.51-Y-1, May 1996.

25. "Space Communications Protocol Standards (SCPS) Space Technology Research Vehicle (STRV) Consolidated Final Test Report," Technical Planning Report SCPS-D71.340-Y-1, September 1996.

26. Robert C. Durst, Darrell E. Ernst, Gregory J. Miller, and Eric J. Travis, "Final Report of the SCPS-TP Testing on the UK DRA STRV," MITRE Technical Report MTR 96W0000075V01, September 1996.

APPENDIX I

DERIVATION OF OPTIMUM PACKET LENGTH FOR ARQ WIRELESS DATA COMMUNICATIONS[1]

1. Introduction

Consider a packet-switched data network containing two nodes on the ground communicating over a full-duplex, wireless link through a (surrogate) satellite relay. We assume for efficiency that a go-back-N data link protocol[2] with automatic repeat request (ARQ) is used in which packets are transmitted continuously in the forward direction, and an error detection scheme is employed with negative acknowledgments (NACKs) contained in the headers of packets sent in the return direction (e.g., [1], pp.121–124). When a NACK is received for a given packet, that packet and the $N - 1$ succeeding packets are retransmitted.

Because the link is wireless, it is characterized by a bit error rate (BER), which may be stochastic or time-varying. If the packet length is small in relation to the size of the header, the link is operating inefficiently since the overhead is too high. At the other extreme, if the packet length is too large for the given BER, packets are likely to be received in error requiring numerous retransmissions and resulting in reduced throughput. Therefore, there is an optimum packet length for which the data throughput is maximized.

2. Analysis

We adopt the following nomenclature:

d = is the altitude of the relay satellite in meters
$c = 3 \times 10^8$ m/sec is the speed of light
$T_r = 2d/c$ is the one-way ground-to-ground propagation time through the satellite relay in seconds
R is the satellite channel uplink/downlink data rate in bits/sec
L is the length of the packet (including the header) in bits
M is the length of the packet header in bits
$T_p = L/R$ is the time required to transmit one packet over the channel in seconds
ε is the bit error rate (BER, the probability that a given bit is received in error)

We assume that the network contains bit error detection but not correction capability, so that a packet needs to be retransmitted if any of its L bits are received in error. Therefore, the packet error or retransmission rate is given by

$$p = 1 - (1 - \varepsilon)^L \approx L\varepsilon ; L\varepsilon << 1 \qquad (1)$$

[1] This analysis mirrors that of Schwartz on pp. 119-133 of [1].

[2] The high-level data link control (HDLC) protocol developed by the International Standards Organization (ISO) is a particular implementation of the go-back-N class.

With probability 1–*p*, a given packet is successfully transmitted on the first try, and the time required to transmit that packet is then trivially T_p. With probability $p(1–p)$, the first try is unsuccessful but the packet is successfully transmitted on the second attempt. Assuming that a NACK is returned as a result of the initial unsuccessful transmission embedded in the header of a return link packet of *L* bits, and using a go-back-*N* transmission protocol with an ARQ scheme, the total transmission time for this packet is $3T_p + 2T_r$ (i.e., the delay introduced due to the single acknowledgment-retransmission sequence is $2T_p + 2T_r$). Continuing with this recursive line of reasoning, the average (expected) time to successfully transmit a given packet is

$$\begin{aligned} T_{avg} &= (1-p)T_p + p(1-p)(3T_p + 2T_r) + p^2(1-p)(5T_p + 4T_r) + \ldots \\ &= T_p + 2(T_p + T_r)(1-p)(p + 2p^2 + 3p^3 + 4p^4 + \ldots) \\ &= T_p + 2(T_p + T_r)(p + p^2 + p^3 + p^4 + \ldots) \\ &= \frac{(1-p)\,T_p + 2pT_r}{1-p} \end{aligned} \tag{2}$$

Then, because each packet contains $L - M$ data bits, the normalized data throughput can be expressed as

$$\rho \equiv \left(\frac{L-M}{L}\right)\frac{T_p}{T_{avg}} = \left(1-\frac{M}{L}\right)\frac{1-p}{1+\alpha p} \rightarrow 1-\frac{M}{L};\; p \rightarrow 0 \tag{3}$$

where, for convenience, we have introduced the parameter

$$\alpha \equiv 1+\frac{2T_r}{T_p} = 1+\frac{4dR}{cL} \tag{4}$$

3. Example

Consider a hypothetical example in which each packet contains a 5 byte (octet) header, as in the Asynchronous Transfer Mode (ATM) protocol, so that $M = 40$ bits. Furthermore, assume a geosynchronous earth-orbiting (GEO) relay satellite, with $d = 3.6 \times 10^7$ m and $T_r = 0.24$ sec. Finally, suppose that the wireless ground-to-ground data link is a T1 channel with $R = 1.544$ Mbits/sec. Then Equations 1, 3, and 4 specify the normalized data throughput ρ as a function of the packet length *L* for a given BER ε. This performance measure is plotted in Figure 1 below.

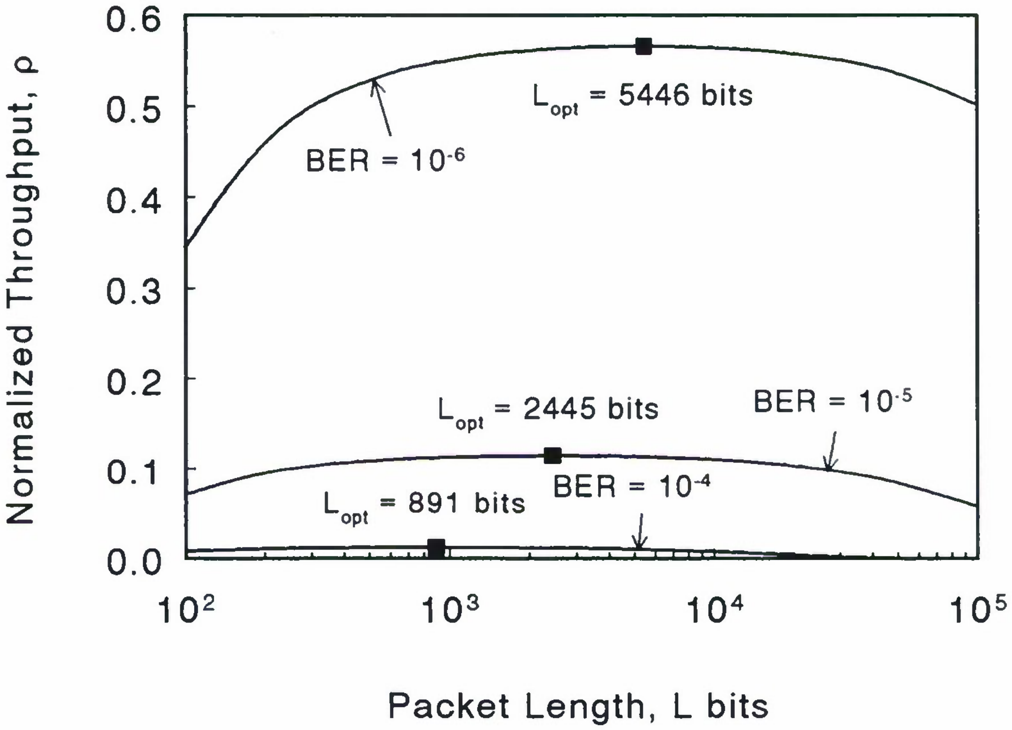


Figure 1. Optimum packet length, L_{opt}, that maximizes the normalized data throughput for a wireless, ARQ, ground-to-ground, T1 (R = 1.544 Mbits/sec) data link through a geosynchronous satellite relay using packets with an M = 40 bit header.

4. Conclusions

It is evident that for each BER, there is an optimum packet length, L_{opt}, which decreases monotonically as the BER increases. However, the throughput performance degrades very gracefully for suboptimal packet lengths $L \neq L_{opt}$. Also, as might be expected, the maximum throughput increases with decreasing BER, approaching 1 (since $L_{opt} >> M$) as the BER approaches zero.

Note that ATM was designed for high-speed transmission over high-quality fiber optic communication channels with nominal BERs of 10^{-9} or better. Although the performance for this BER is not shown in Figure 1, the maximum throughput is ρ_{max} = .998694 at an optimum packet length L_{opt} = 141,000 bits. By comparison, ATM packets (cells) are fixed at 53 bytes (L = 424 bits), for which the throughput degrades to .904989 or 90.6% of its maximum value at BER = 10^{-9}, which again illustrates how robust the throughput performance is as a function of the packet length.

Reference

1. Mischa Schwartz, *Telecommunication Networks: Protocols, Modeling and Analysis*, Addison-Wesley, Reading, Massachusetts, November 1987.

APPENDIX J

UPLINK AND DOWNLINK FREQUENCY ASSIGNMENTS AND REUSE

A realistic frequency allocation for the Warfighter's Internet is probably in the range of a few MHz to 10s MHz total for uplink and downlink. Maximizing the throughput available to terminals using this limited frequency resource can be achieved by reusing as often as possible within the theater the available uplink and downlink frequencies (see Figure 1).

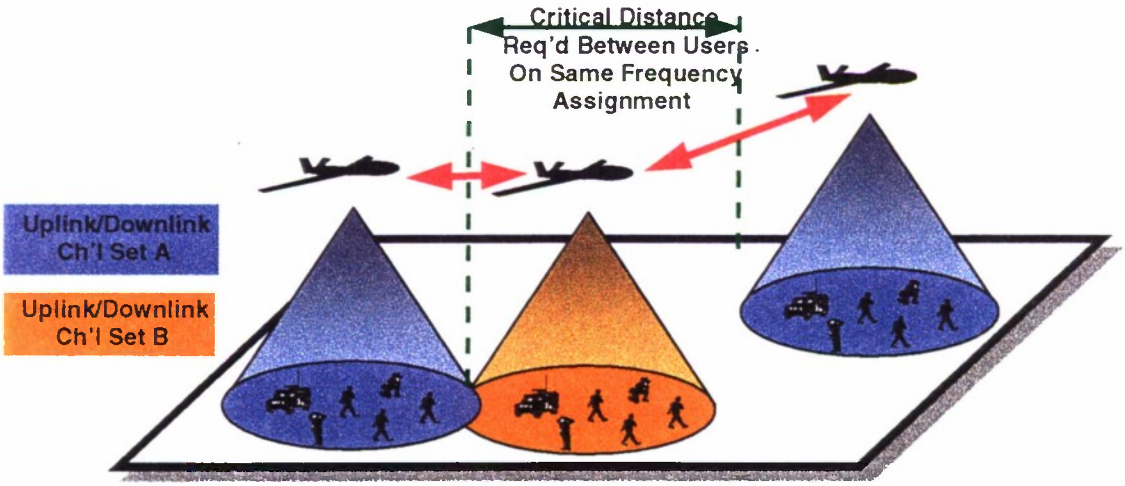


Figure 1. Warfighter's Internet frequency/channel reuse concept.

Depending upon whether orthogonal or non-orthogonal modulation techniques are used, either no interfering users or only a few interfering users can be allowed in any one service area. Additional terminals assigned to the same time/frequency channels must be physically separated from terminals using those channels by some critical distance. Without this critical distance, terminals uplinking to one node would potentially interfere with a different aerial node, or aerial nodes downlinking to terminals might interfere with signals coming from a different aerial node.

Many different algorithms are possible for assigning uplink and downlink channels based upon a critical separation distance. These algorithms can be defined in terms of the nominal coverage area for each aerial node, and the linear ground distance required between terminals assigned to the same uplink and/or downlink time/frequency assignment.

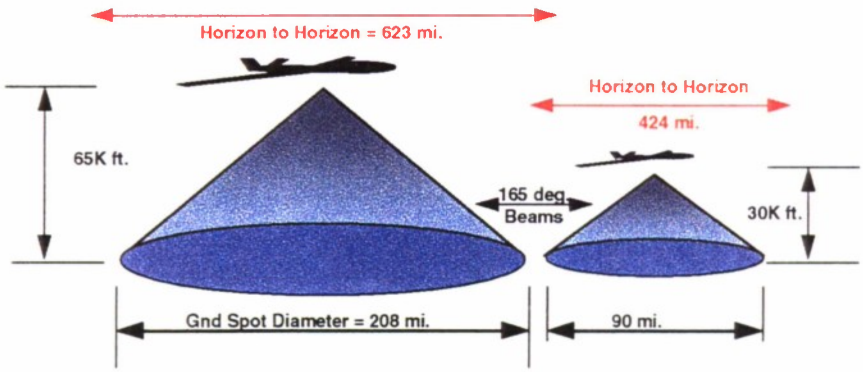


Figure 2. Range of nominal coverage scenarios for handheld service.

Figure 2 above shows the nominal coverage area on the ground for the range of platforms being considered for the aerial backbone. The critical distance of separation required between terminals or groups of terminals reusing channel assignments is generally less than the horizon-to-horizon distance shown above, but will vary from scenario to scenario.

Algorithms for reusing time/frequency allocations fall into two primary categories:

Complete Knowledge Algorithms:
Those algorithms requiring all aerial nodes know the GPS location and channel assignments of each terminal, and update this database on a continuous basis, negotiating as required new assignments.

Incomplete Knowledge Algorithms:
Those algorithms not requiring a well connected backbone and therefore not requiring detailed terminal location and channel assignments be known by all the aerial nodes.

Complete Knowledge Algorithms can be described fairly simply by using Figure 3.

A terminal is being served by the aerial node in the center of the diagram. Given the terminal's GPS location within the nominal coverage area of one aerial node, the same channel assignments cannot be made for a specific critical distance from the terminal, otherwise they may interfere with the aerial node while it is still serving the original user.

Continuing to assign additional users the same channel assignments as above, we get users optimally packed on the hex grid below, each point the same critical distance from each other (see Figure 4).

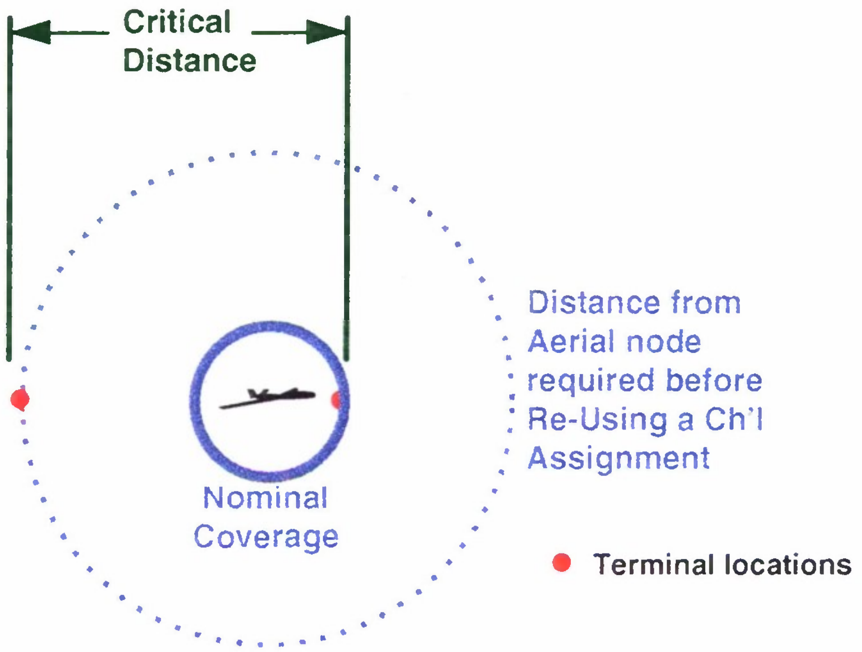

Critical Distance
Distance from Aerial node required before Re-Using a Ch'l Assignment
Nominal Coverage
● Terminal locations

Figure 3. Potentially interfering terminals.

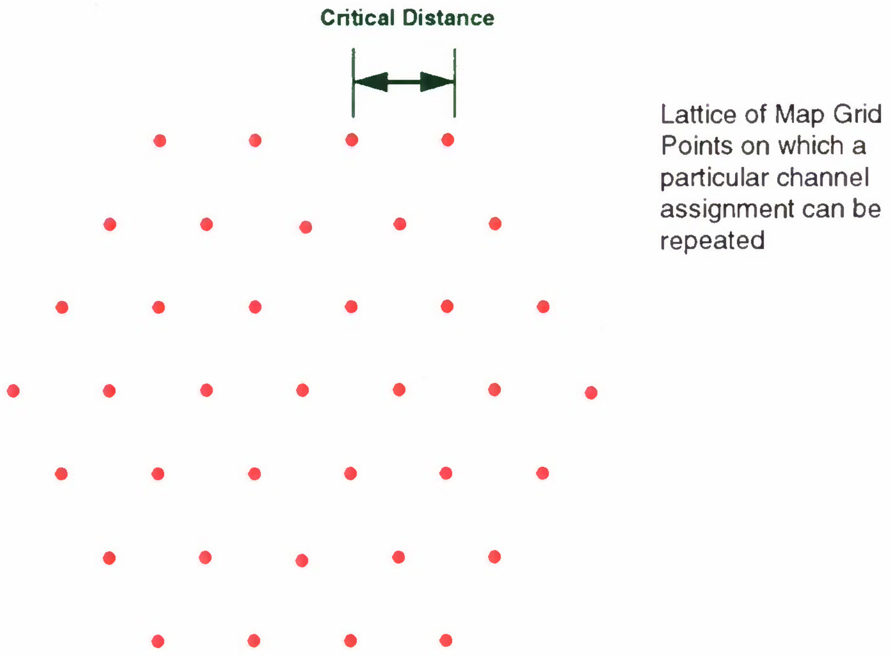

Critical Distance
Lattice of Map Grid Points on which a particular channel assignment can be repeated

Figure 4. Hex grid for reuse of channel assignments when complete knowledge known.

In practice this packing density would never be achieved since users would come and go at GPS locations other than the ones above. The exact placement of the above lattice on the map is fixed by the first user assigned to a specific channel. Each time a new user requested service he would be checked against all other terminals' GPS locations and channel assignments in the theater-wide database before assigning a channel according to some algorithm. This type of algorithm does, however, allow all channels to be assigned in an arbitrarily small area on the map, and this density of assignments to be repeated throughout the theater.

Incomplete Knowledge Algorithms can only assume knowledge of a terminal request for service in a particular area. These algorithms reduce to pre-assigning frequency channels to areas on the map and using standard N-Way reuse techniques to repeat the assignments at critical distances from one another.

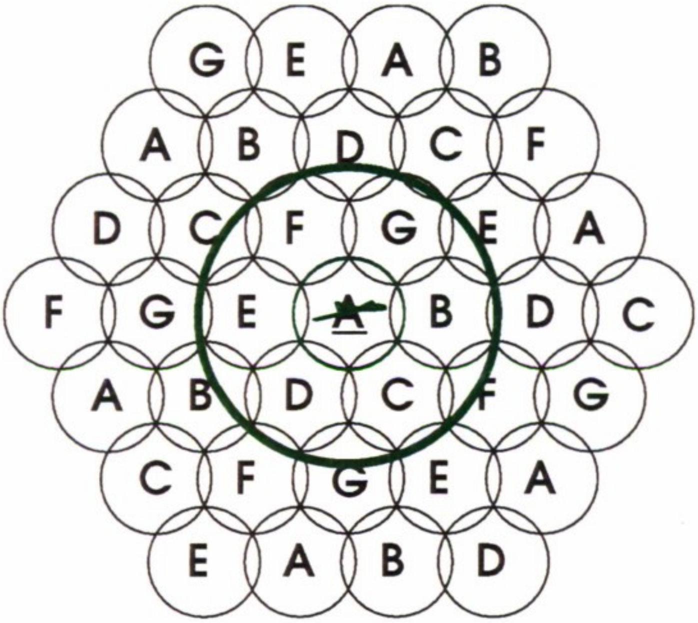


Figure 5. Example of 7-way reuse.

The example given in Figure 5 above shows 7-way reuse. At the center is a nominal service area A using nominally 1/7 of the channel resources (assuming they are divided up ahead of time in a uniform manner. All terminals and all aerial nodes know the pattern on the map and therefore know when to transition from one 'cell' to another. If the aerial nodes are not connected, then it is up to the terminals to manage hangovers when aerial nodes or terminals move. (Note that each terminal will know with which aerial node(s) it can communicate.) The resultant efficiency of reuse in this case is worse than 7 times less than the Complete Knowledge

Case. In this scenario, the lack of complete information forces the critical distances required to separate service areas to be larger than those required to separate individual users.

In practice, the Warfigher's Internet will require resource assignment algorithms which are a hybrid of the two schemes above. Since all the users are highly mobile, reusing channels at exactly the optimum distance forces frequent handovers. So real world algorithms will allow for a 'fuzz' factor associated with the location of a user and thereby take on aspects of both the algorithms outlined above.

In summary, it can be seen that there are many advantages to aggressive frequency reuse in The Warfighter's Internet, but the full exploitation of these advantages requires that the frequency reuse algorithms be designed into the theater-wide networking coordination between the nodes of the aerial backbone.

APPENDIX K

WARFIGHTER'S INTERNET FUNCTIONAL SPECIFICATION FOR ARMY TACTICAL COMMUNICATIONS

1. Introduction/Overview

The Warfighter's Internet will provide a robust, self organizing airborne communications infrastructure to tie together the joint fighting force of the future. The Warfighter's Internet will provide communications coverage from the individual dismounted subscriber to the command posts. The increase in horizontal and vertical communications will enhance commander awareness and aid in all phases of operations from battle planning to engagement. The Warfighter's Internet will blanket the battle space tying together the capabilities of the Army with the joint force structure.

2. Operational Need and Concept

2.1 Army Scenarios

The pace at which the battle line can advance with today's fighting technologies is over-extending the existing communications infrastructure. Developing a communications infrastructure to support such a highly mobile, spatially diversified, and information hungry fighting force is the engineering tactical dilemma of the year 2000 and beyond. Currently, deployed line-of-sight (LOS) ground-based communications systems have their limitations in providing the fighting solider the necessary tools to reduce the battle space dilemma. As the Army moves ahead in its digitization efforts, dispersed connectivity should be meshed together to provide a fully connected mobile force. Since ground-based communications have inherent limitations and satellite resources are limited in all frequency bands, the Army should develop an airborne Warfighter's Internet to tie together the Army and joint forces into a single fighting force.

Forward deployed scouts, small unit operations (SOU), and sensor platforms are three differing Army scenarios that could utilize a Warfighter's Internet (WI) to improve mission effectiveness. The WI can support Army forward-deployed users by providing reliable, high data rate connections to command operations centers in support of mission objectives. The types of traffic required to support these scenarios is voice, bursty data communications (sensor information, position reports, spot reports, etc.), still imagery file transfers, and limited video connections. These forward-deployed units would be communicating to the rear to a concentrator node like a Radio Access Point (RAP). The RAP could provide an interface so that forward deployed users will have a reach-back capability into the command groups. The RAP can also provide the capability to integrate the joint services in meeting mission requirements.

Echelon-to-echelon traffic flow at key user positions is restricted using the currently in place ground-based communications infrastructure. The WI could provide a select number of users at key mission locations the capability to communicate across the battlefield both horizontally and vertically without the burden of the limited communications capabilities of today's ground mobile radios. Allowing the commander to maintain communications with the key components across the spatially dispersed battlefield, without restrictions due to the mobility of the ground-based communications infrastructure, will greatly enhance the commander's effectiveness.

Concentrator nodes like the RAP will play a key role in tying together the Army communications and information infrastructure. Integrating the communications infrastructure of a highly mobile platform like a RAP is essential to battle effectiveness. Concentrator nodes that tie together the Army's logistics, fire support, command information, situational awareness, and Army battle space picture should be fully connected horizontally and vertically at all times. The WI could provide the beyond-line-of-sight (BLOS) communications assets for concentrator nodes like the RAP to meet mission requirements.

For the WI to adequately support the Army/Joint service needs, it should provide reliable communications between small groups of dispersed users, echelon-to-echelon battle planning traffic, and information exchange between all these groups through concentrator nodes like the RAP. Over the WI, the needed command and control traffic, sensor information, requests for medical, logistics and fire support services can be requested and responded to effectively.

2.2 Forward Deployed Units/Scouts

The WI should provide connectivity to forward-deployed units (extended beyond LOS communications support) and scouts collecting/relaying intelligence information. A point of insertion into the mainstream Army network could be a concentrator node like the RAP. The WI should be able to provide connectivity to a RAP and forward-deployed units simultaneously.

2.2.1 Traffic Performance Requirements

The traffic performance requirements outlined in Table 1 assume that all voice transfers are packet voice and the subscriber terminal interfaces and/or functionality is to the current Appliqué. Voice and SA data speed of service (SOS) and quality of service (QOS) requirements vary. The WI will prioritize the establishment of resources for voice connections between the originator and the destination subscribers. If the destination subscriber for a voice call is unreachable, immediate notification should be provided to the originating subscriber. Still images and BC data SOS requirements are moderate and QOS requirements are high. The WI should provide the subscriber access to the network with minimal delay. A voice call setup should occur in less than 1 second, and still images should be transmitted from scout platforms to a destination terminal within 30 seconds for image file sizes of 30 kbytes.

TABLE 1
Traffic Performance Requirements

Traffic Type	Speed of Service	Quality of Service
Voice	High; with low delay	Medium; low PER
Still Images	Moderate	High; low PER
SA Data	High	Moderate
BC Data	Moderate	Guaranteed delivery

***NOTE: How will the WI support SA? Will the WI support call for fires and engagement operations?*

2.2.2 Mobility

The WI will provide reliable services to support dismounted soldiers and vehicle mounted systems. The WI will provide sufficient area of coverage and meet the traffic performance requirements in Section 2.2.1 to support a dismounted soldier moving at roughly 3 mph and a vehicular-mounted system moving at roughly 30 mph. The dismounted soldiers and vehicular mounted systems will employ omnidirectional antennas.

Addressing, routing, and forwarding of all subscriber traffic will be the responsibility of the WI. The WI will resolve mobile addressing, provide store and forward capabilities, and route packets without subscriber intervention.

2.2.3 Area of Coverage

The WI should provide coverage from the forward-deployed unit to WI subscriber link at a Tactical Operations Center (TOC) further to the rear. Since a scout or forward-deployed unit will be communicating to the rear, the area of coverage specified (see Figure 1) will represent the maximum vertical range of coverage the WI should provide to the Army user community. The vertical range of coverage needed between a TOC and a forward deployed unit/scout is approximately 100 km.

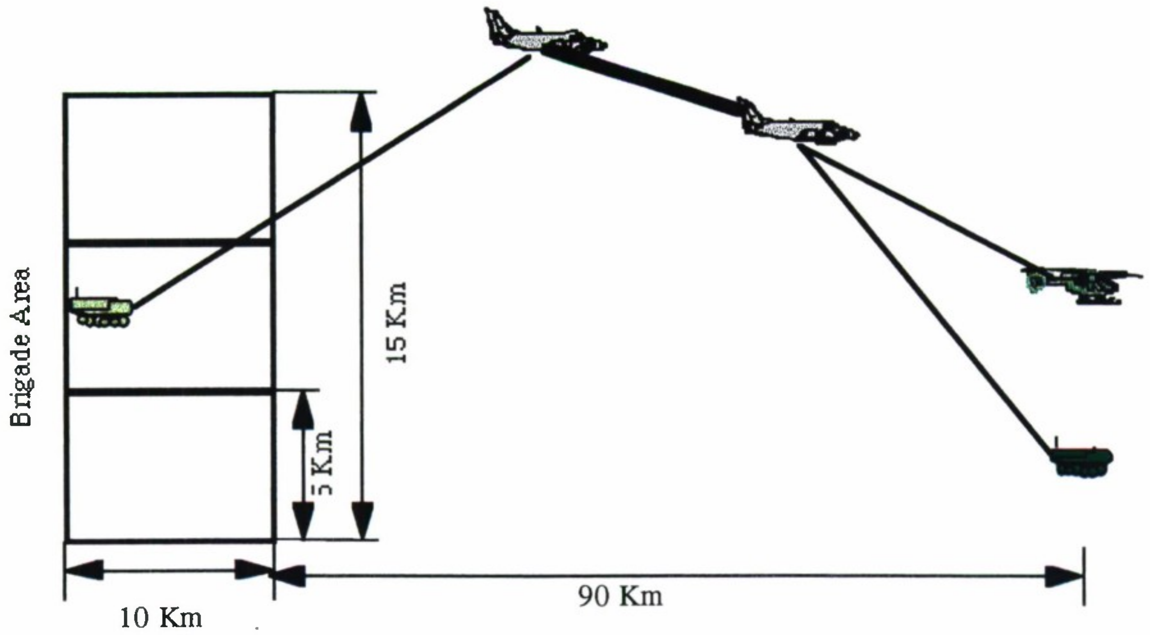


Figure 1. Scout area of coverage.

***Note: Scouts and forward deployed units may be on the fringe of coverage for the WI. Levels/grades of service, SOS, QOS, and area of coverage analysis should be performed to insure that the dismounted and forward-deployed user requirements are satisfied, while maintaining connectivity to units deployed to the rear.*

2.3 Echelon to Echelon

Echelon-to-echelon communications in this context refers to the horizontal/vertical communications between key nodes across the battle space that require a more direct line of communications to facilitate battle planning/execution. The WI could provide a reliable communications path between command groups independent of terrain or geographic position.

2.3.1 Traffic Performance Requirements

The traffic performance requirements are the same as in Section 2.2.1 with the following addition. Current Army field exercises have demonstrated to the commanders the added value of Video Teleconferencing (VTC). To properly support battle planning, a limited video capability should be supported by the WI to enhance battle planning. The WI should provide the capability to transmit and receive real time video point-to-point and provide half-duplex broadcasting for conferences involving more than 2 members. The WI should support to all users a minimum capability of 8 frame per second video/voice.

***Note: The actual numbers of video frames per second supported by the WI should be evaluated based on its impact on other WI subscribers. An analysis should be performed to* *determine the optimal number of frames per second, use of ground based directional antennas and so forth to achieve a video capability.*

2.3.2 Mobility

The mobility requirements are the same as outlined in Section 2.2.

2.3.3 Area of Coverage

The maximum vertical range of coverage is as specified in Section 2.2. To determine horizontal range of coverage it will be assumed that a defensive scenario for a Brigade-sized deployment is roughly 150 sq. km (10 km deep by 5 km wide). To provide echelon-to-echelon communications, assuming the brigades will be dispersed across the battle space, the WI will provide a horizontal range of coverage of 50 km. The total area of coverage that should be provided by the WI to meet echelon-to-echelon and forward unit/scout reliable communications is 5000 sq. km (100 km deep by 50 km wide).

2.4 RAP/JMCOM/Concentrator Nodes

The following services are required by the WI as a minimum to support RAP/JMCOM/ concentrator node connectivity referred to as RAP service throughout the remainder of this document. RAP services can be evaluated as two separate options; segregated service or integrated service. WI segregated RAP service will allow ATM cells to be transmitted to the WI, cell switched over the backbone architecture, and broadcast down as ATM cells to a destination RAP. Segregated service will not provide the capability for ATM traffic to be forwarded to individual subscriber terminals supported by the WI. WI integrated service will allow ATM cells to be transmitted to the WI, cell switched over the backbone architecture, and broadcast down as ATM cells to the destination RAP. It will also provide the capability for IP subscribers connected to a RAP to communicate directly with individual subscribers supported by the WI. There are several methods that can be used to provide an integrated service, the methods are detailed in Section 3.1.3 RAP Services, of this functional specification.

***Note: The Army, Navy, and Air Force concentrator node services (RAP/JMCOM) need to be further integrated into this document. This appendix is an initial cut; future efforts should tie together the joint service interoperability. .*

2.4.1 Traffic Performance Requirements

The RAPs high bandwidth connection is Asynchronous Transfer Mode (ATM) based. The RAP will act as a concentrator node supporting VTCs, multimedia applications, voice, mission data, and subscriber traffic. The WI should provide a sufficiently low cell error rate to support point-to-point wireless ATM transmissions. The WI uplink, backbone, and downlink architecture should maintain a composite cell error rate less than 0.01% for all RAP links.

2.4.2 Mobility

The WI should be able to support a RAP moving at roughly 30 mph and a communication rate not less than 1.5 Mbps. The RAP should provide a tracking antenna to maximize link connectivity. The mobile platform will attempt to maintain link connectivity to a single ACN to minimize communications link outages and the impact on WI mobility management.

Cell switching over the WI backbone between RAP platforms will be the responsibility of the WI. The WI backbone architecture will be capable of resolving ATM addressing and providing the QOS/SOS defined by the ATM circuit.

2.4.3 Signalling Channel

Currently the RAP concept utilizes a signalling channel over the NTDR to discover and affiliate with other RAPs. The WI must provide a signalling channel for RAPs to establish connectivity with the WI network.

2.4.4 Area of Coverage

The area of coverage requirements in Section 2.3.3 defines the area of coverage that a WI should provide to support Army forces. When integrating the joint community into the area of coverage requirements, the area of coverage is greatly increased (see Figure 2). For inter-communications between Navy systems like JMCOMs and Army systems like the RAP joint discussions must be conducted.

3. Major Functions and Interfaces

The WI for functional purposes will be broken down into four major subcategories:

- Data Transport Services
- Intra-Airborne node
- Inter-Airborne node (backbone)
- Network management/mobility

3.1 Data Transport Services

There are two primary types of data transport services being provided by the WI. The WI will provide an IP packet voice, data, and still imagery service to small mobile platforms, dismounted soldiers, and small operation units. The WI will also provide wideband data link services to concentrator platforms like the RAP and be capable of switching ATM cells. From this point forth, individual subscriber (IS) service will denote the IP data/voice channel and RAP will denote the composite uplink from a concentrator platform.

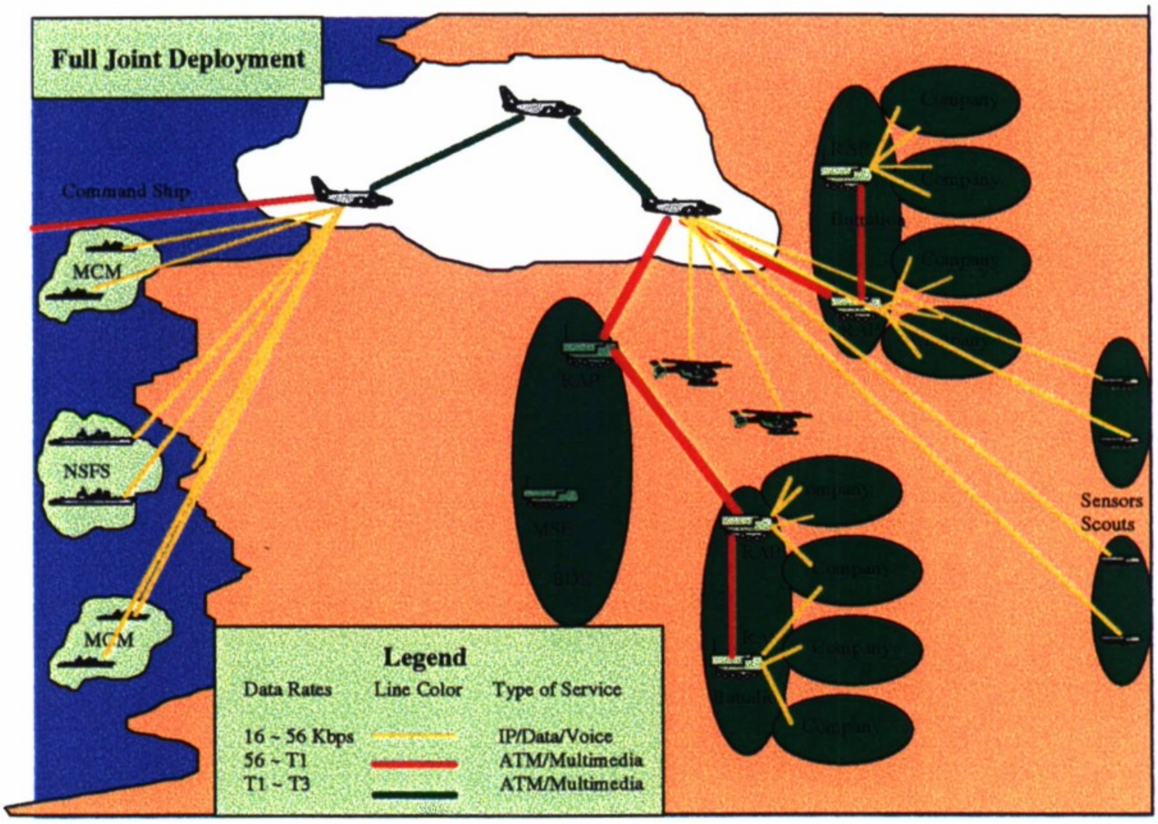


Figure 2. Joint deployment communications coverage.

3.2 Individual Subscriber Services

Subscriber services will include IP packet voice, IP data, and still imagery. The WI will be capable of supporting a varied Quality of Services (QOS) and Speed of Service (SOS) requirements to meet the user mission goals. Subscriber terminals will automatically register with an Airborne node within the WI. The WI will provide routing and mobility management necessary for the desired communications. The WI will provide subscribers low data rate packet voice (2.4 - 4.8 kbps) and higher data rates for data/still imagery (16 - 56 kbps). The WI will provide subscriber services for up to 1000 subscribers (30 in initial demonstration) per airborne node.

***Note: An analysis should be performed to determine the maximum number of subscribers that can be supported realistically taking into account such factors as ACN weight restrictions, power requirements, total usable bandwidth, percentage of users off hook and so forth.*

3.2.1 Subscriber Uplink

The subscriber uplink channel will be accessed through a lightweight handheld unit. The unit will be capable of supporting multiple uplink waveform techniques as defined by the WI analysis/study team. Each waveform will be designed/implemented to provide the maximum data rate/reliability for the user service it supports. The uplink and downlink channels will have asynchronous data rates to take advantage of a downlink broadcast capability.

For BC data, data link control (DLC) will be implemented into the subscriber uplink to provide for error/flow control. DLC options include connection-oriented (CO) and connection-less (CL) control. The subscriber uplink channel must select its DLC to efficiently support the varied types of packet sizes, message lengths, and message priority.

3.2.2 Subscriber Downlink

The subscriber downlink channel will be a shared broadcast channel from the airborne node to the subscriber terminals. The subscriber downlink terminal will be capable of receiving broadcast/multicast based on IP and MAC layer addressing. One or more channels shall support individual subscribers and RAP users, respectively.

3.2.3 Signalling Channel

An out-of-band signalling channel between the subscriber and the airborne nodes is needed. This channel will be used to allocate ACN receiver/transmitter resources as required to meet the subscriber requirements, register subscribers with WI mobility management, provide link quality monitoring interaction, and conduct ACN hand-off procedures.

3. 3 RAP/JMCOM/Concentrator Node Service

As addressed in Section 2, RAP services have two options: integrated service or segregated service. WI segregated RAP service will allow ATM cells to be transmitted to the WI, cell switched over the backbone architecture, and broadcast down as ATM cells to the destination RAP. Segregated service will not provide the capability for ATM traffic to be forwarded to individual subscriber terminals supported by the WI. WI integrated service will allow ATM cells to be transmitted to the WI, cell switched over the backbone architecture, and broadcast down as ATM cells to the destination RAP. It will also provide the capability for IP subscribers connected to a RAP to communicate directly with subscribers supported by the WI.

3.3.1 Segregated Service

The segregated individual subscriber (IS)/RAP service is depicted in Figure 3. The WI will provide the capability to perform RAP-to-RAP switching for beyond-line-of-sight relay capability. The WI will provide 1.5 Mbps, full-duplex circuits for RAP communications. RAPs will provide tracking antennas in an attempt to maintain connectivity with a single airborne node. The RAP will be responsible for providing routing amongst its IP subscribers, but the WI will provide necessary switching and address mapping/resolution between RAPs. It is the responsibly of the RAP to perform subscriber discovery with other RAP platforms. The WI will perform cell switching to RAP platforms.

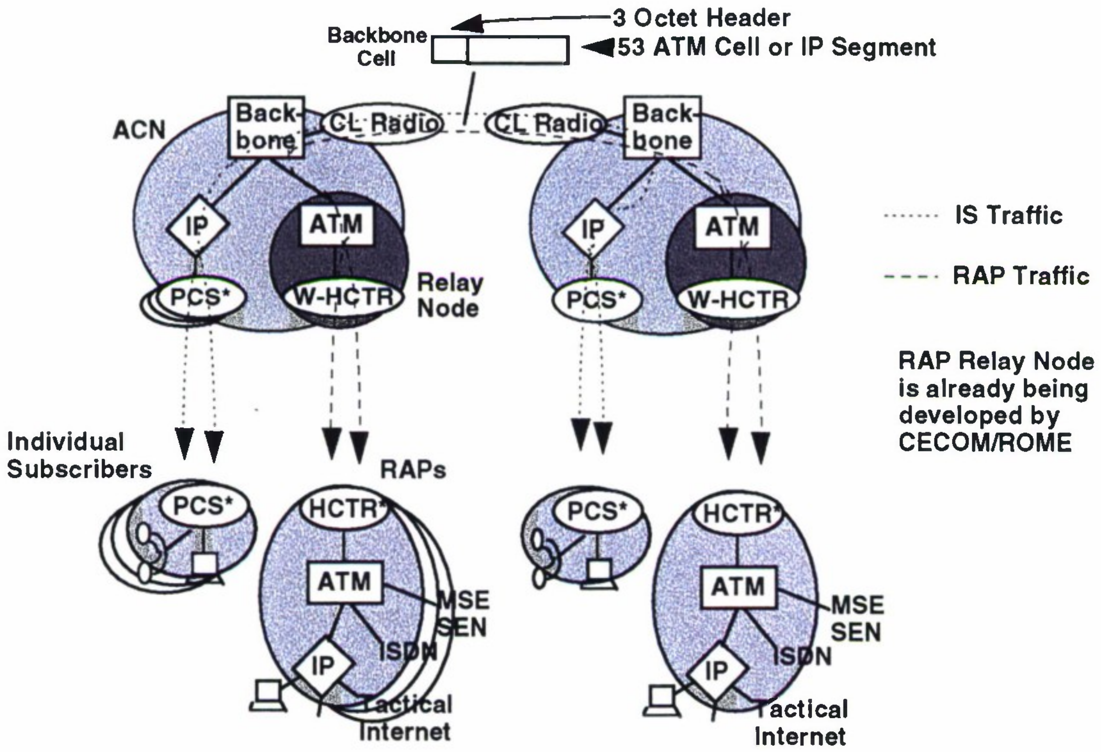


Figure 3. Segregated RAP architecture.

Under the segregated services option, RAP IP subscribers will not have access to subscribers supported by the WI. This option, however, minimizes WI/subscriber mobility management complexity.

***NOTE: The reader should note that the PCS* and HCTR* radios are called out for clarity of design but are not intended to recommend a solution. Also note, currently the RAP will utilize the NTDR for discovering other RAPs across the battlefield. The WI will have to develop a method concurrently with the RAP/JMCOMS/concentrator node developers to resolve the discovery of the WI from ground terminals.*

3.3.2 Integrated Service

As other services begin to develop/field their wideband network traffic systems, the ATM service and integrated ATM/IP subscriber communication will become essential. The Navy, Air Force, and Army must be able to intercommunicate over the WI. The IP subscribers on the tactical Internet, scouts, and forward-deployed units will be able to extend their communications reach across the battle space.

Unlike the segregated service, the WI will be responsible to provide routing of IP subscribers connected to a RAP to other RAP IP subscribers and to individual subscribers supported by the WI. In the first implementation option for integrated service, the RAP will route IP traffic destined for individual subscribers registered on the WI using an IP subscriber radio. The integrated RAP/individual subscriber service can be accomplished through two different architectural designs. Figure 4 depicts architectural design 1 and Figure 5 depicts architectural design 2. The major difference between the two architectural designs is in Figure 4 the IP subscriber interface occurs at the RAP, in Figure 5 the IP subscriber interface occurs within the WI ACN.

3.3.2.1 IP Subscriber Interface in Ground Terminal

The WI with integrated service will provide the capability to perform RAP-to-RAP switching for beyond-line-of-sight relay capability. The WI will provide 1.5 Mbps, full-duplex circuit, for RAP communications. A single radio with frequency division multiple access (FDMA) will be utilized for trunked links with airborne platforms. RAPs will provide tracking antennas in an attempt to maintain connectivity with a single airborne node. The WI backbone will determine routes to other registered RAPs and provide addressing mapping/resolution as required. The trunked ATM traffic will be cell switched over the WI backbone to the destination RAP. RAP will aggregate addresses of their IP subscribers to reduce the addressing and discovery overhead for mobility management. The RAP will route IP subscriber traffic destined for other RAP IP subscribers over its ATM link. It is the responsibility of the RAP to perform subscriber discovery within other RAP platforms not supported by the WI.

Figure 4 depicts the RAP IP subscriber interface to the WI. Note that there are two interfaces to the WI from the RAP, one over the ATM trunk and the other over a subscriber radio.

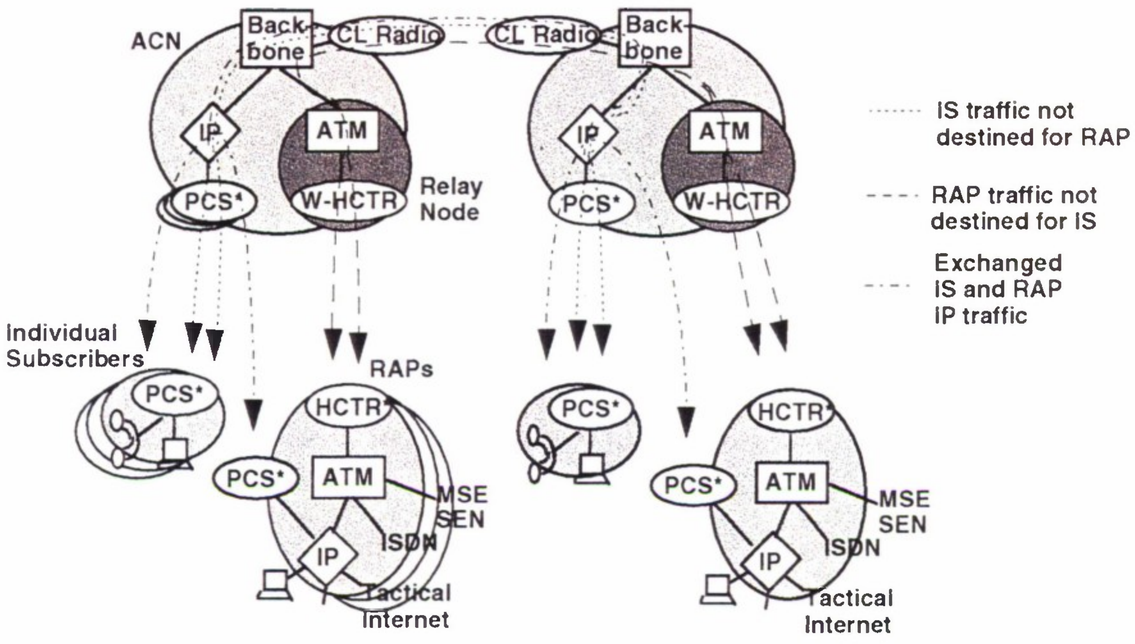


Figure 4. RAP IP subscriber interface architecture.

***Note: The reader should note that the PCS* and HCTR* radios are called out for clarity of design but are not intended to recommend a solution. Currently the RAP will utilize an NTDR-like radio for discovering other RAPs across the battlefield In this approach, the WI subscriber terminal may be used for RAP WI discovery.*

3.3.2.2 IP Subscriber Interface in Airborne Platform

In the second implementation option for integrated service, ATM cells transporting IP traffic destined individual subscribers will be transported as ATM cells to the ACN router of their destination individual subscriber and reassembled into IP for downlink delivery. The opposite procedure will apply for individual subscriber to RAP IP subscriber communications. ATM cells transporting traffic for other RAPs will be cell switched by the WI and routed over the backbone as trunked traffic. The WI backbone will determine routes between RAPs and provide addressing mapping/resolution as required. The RAP will be responsible for providing its IP subscriber addresses (aggregated) to WI mobility management.

Figure 5 depicts the WI IP subscriber interface to the RAP platforms. Unlike implementation option 1, this option avoids the need for a separate subscriber terminal at RAPs but requires an NTDR-like radio at ACNs for RAP ACN discovery.

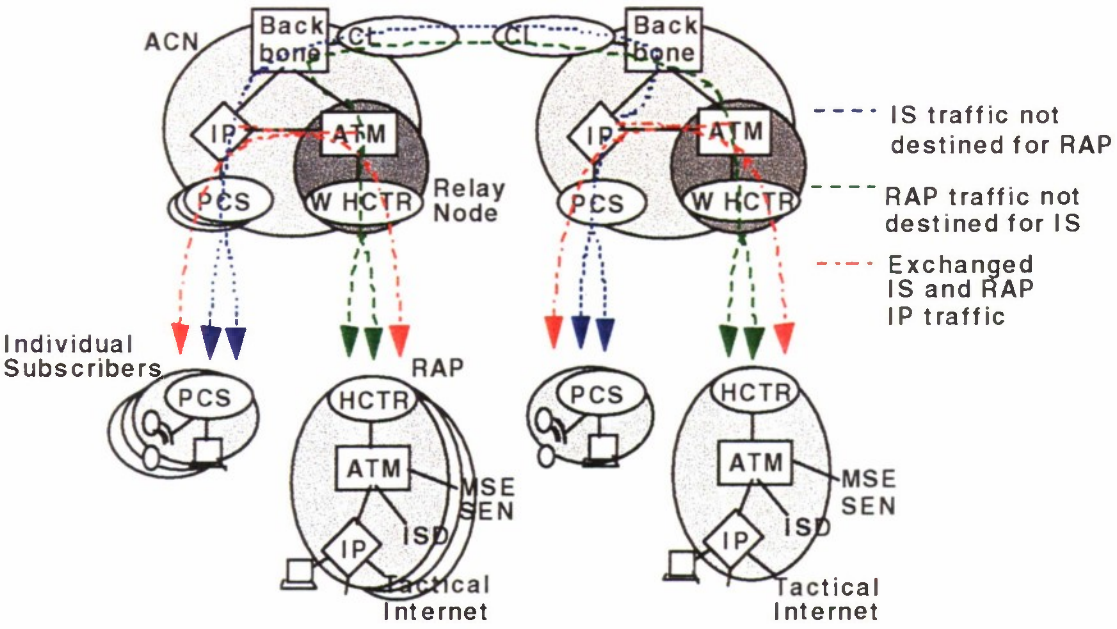


Figure 5. WI IP subscriber interface architecture.

**Note: Mobility management of the digital battlefield is not a small task to undertake. If several organizations are responsible to develop their own mobility managers and then integrate their solutions into the WI, the outcome could be as hard to resolve as the mobility issue itself. Consequently, it is recommended that the WI take on all of the mobility management responsibilities for systems interconnected to the WI.*

3.4 Intra-Airborne Node

The airborne node will provide an IP routing capability to support subscribers registered with that airborne node. The airborne node will provide the capability to forward subscriber traffic to another airborne node supporting the destination subscriber. The airborne node will provide any data translation that may be required to support the implemented inter-airborne node architecture. Level of service, QOS, and SOS will be guaranteed and continuously monitored by the airborne node. The airborne node will be capable of determining a new route for traffic to minimize the impact on service.

The airborne node will be capable of supporting resource reservation for individual subscriber voice traffic and ATM trunked traffic.

3.5 Inter-Airborne Node

Data and voice traffic will be statistically cell multiplexed in optional broadcast or point-to-point cross-links between neighboring airborne nodes. Backbone routing/switching will be used for directing traffic over backbone cross-links from source to destinations. Routing/ switching will provide an efficient path selection and limit overhead for uni- and multi-casting and for conditions of moderate backbone dynamics and varied link qualities.

3.5.1 Signalling Channel

The inter-airborne node will provide a low data rate signalling channel between airborne nodes over an omni antenna. The signalling channel will be used to allow for graceful insertion and planned exit of airborne nodes. The link will be capable of supporting unplanned exits of an airborne terminal without impacting traffic that was not supported over the failed link. The backbone will be scalable to accommodate multiple backbone nodes. Mobility management routing overhead will be supported over the signalling channel.

One option for the signalling channel is to require at least one transmitter per airborne node. It may also require n – 1 receivers per airborne platform, where n equals the number of backbone nodes. Each backbone node will require a unique and independent transmit frequency to allow full connectivity. The scalability of the signalling channels will be dependent on the number of independent transmit frequencies that can be provided to the WI on a non-interfering basis.

3.5.2 High Bandwidth Data Links

High data rate backbone link will support MUXed individual subscriber and RAP ATM traffic. Resource reservation will be set-up by the signalling channel. The backbone architecture will accommodate the QOS and SOS defined by the signalling channel. The backbone will utilize multicast capabilities to reduce traffic over the backbone. The link will provide both virtual-circuit switching and datagram switching.

The high data rate channel will require at least one transmitter/receiver per directional antenna. The directional link will be established as a point-to-point link. The link should support quick setup and tear down to support the dynamics of the airborne nodes.

3.5.3 Directional Antenna

The high data rate directional antenna will need to be steerable and controlled via the signalling channel. It will provide data throughput greater than 1.5 Mbps. The antenna system will provide a mechanism to facilitate antenna pointing.

3.6 Inter-Airborne Node Management

Airborne node link management will be responsible for UAV node insertion, planned node exit, unplanned exit, antenna control, node congestion, bandwidth reservation, link quality monitoring, radio control, and mobility management (subscribers and backbone nodes).

3.6.1 Node Insertion

The network manager will use the signalling channel to allow nodes to gain access to the network. The manager will continuously monitor for nodes that want to join the network. Once the terminal is allocated channel resources the WI will initiate backbone routing procedures.

3.6.2 Planned Node Exit

The network manager will preplan for routing traffic around a planned node prior to it exiting the backbone. Subscribers that will have effected or reduced support should be notified of WI reduced capability. Every path will be examined to minimize the impact on supported traffic during a planned node exit.

3.6.3 Antenna Control

Antenna control will be provided by the network manager/signalling channel control. The directional antenna will be capable of automatic and manual steering control. The antenna system will have access to any platform orientation information that is available to carry out prescribed commands. The antenna interface should permit the antenna to be directed by specifying the GPS coordinates of both the source and destination nodes.

3.6.4 Node Congestion

During periods of high traffic volume, the network manager will take advantage of different options to reduce node congestion. The options may include some or all of the following:

- Changing the subscriber registration
- Reallocating bandwidth
- Tearing down of reserved circuits
- Modifying routing metrics
- Termination of service by priority

3.6.5 Registration

The WI will provide subscriber registration for subscribers to initially gain WI receive and transmit resource allocations. The subscriber registration is required to support the mobility manager.

3.6.6 Bandwidth Reservation

Bandwidth reservation will be controlled by the network manager. The reservation scheme will support the QOS and SOS outlined for the intra-airborne node. IP traffic reservations will be established using a reservation mechanism like RSVP.

3.6.7 Link Quality Monitoring

The network manager will use the signalling channel to provide the capability to exchange link quality monitoring information. An error detection capability should be integrated into the network manager. Subscribers will use this information to decide to conduct a handover procedure.

3.6.8 Mobility Management for Individual Subscribers

Intra-airborne node mobility will be handled by the network mobility manager. The mobility manager will be able to accommodate subscriber mobility and backbone mobility. The mobility manager will, as a minimum, interface with the inter-airborne node subscriber registration table for each airborne node directly connected. Mobility management traffic will use the signalling channel.

APPENDIX L
TOPOLOGY AND ELECTROMAGNETICS SYSTEMS MANAGEMENT

Radio systems of each US military service share the electromagnetic spectrum with one another, with communications users of the other services and other nations, organizations, and individuals, and with other electromagnetic receiving and radiating systems. Most of these other users rely on access to the spectrum to meet service needs, while some are incidental participants (e.g., welding equipment).

Radio Frequency Network Management is the process by which the services apportion communications resources to most effectively meet mission needs. It is carried out at all levels, including Theater, Force, and Unit. The resources available for apportionment by the manager are the set of assigned frequencies within the spectrum, nodal communications equipment[1] (antennas, radios, cryptographic units, etc.), and relay nodes (including existing and possible future relay pods and palettes, UAVs, balloons, relay towers, etc.). In particular, the Warfighter's Internet consists of a set of nodes that must be managed in this context. For the purposes of this Appendix, "management" is taken to include the planning processes that can take place days or weeks in advance, as well as the management processes that take place in real time. Management includes the decision processes, but does not include the actions taken to execute the process. These actions (e.g., tuning a radio, or executing the setting of other specific equipment parameters) are considered "control" rather than management. The general management and control process relationships are shown in Figure 1. This section deals with the management processes included in the green box. The processes discussed in this section apply to large nodes—command centers, Radio Access Points (RAPs), and ships. For Warfighter's Internet, the process applies to the node(s) performing the Entry Node and Backbone Connection Manager functions.

The communications equipment resources available to a large node in the future will be flexible. The use of software and digital processing will allow the available equipment to be configured into a variable number of "channels." There will be a maximum number of channels, varying by node type. Configurable, modular digital radios are being defined in various forums[2], and being implemented within various programs such as SPEAKeasy (a DARPA-initiated, joint advanced development effort) and the Joint Combat Information Terminal (a US Army implementation program).

[1] In the longer term, equipment may be multi-functional. For example, a receiver could be directed to a navigation, communications, surveillance, or other function, depending on dynamic user requirements. Here we assume the traditional view of dedicated (although very flexible) communications equipment.

[2] Examples include the Programmable Modular Communications System Integrated Product Team (a US joint services definition effort) and the Modular Multifunction Information Transfer System Forum (an international effort).

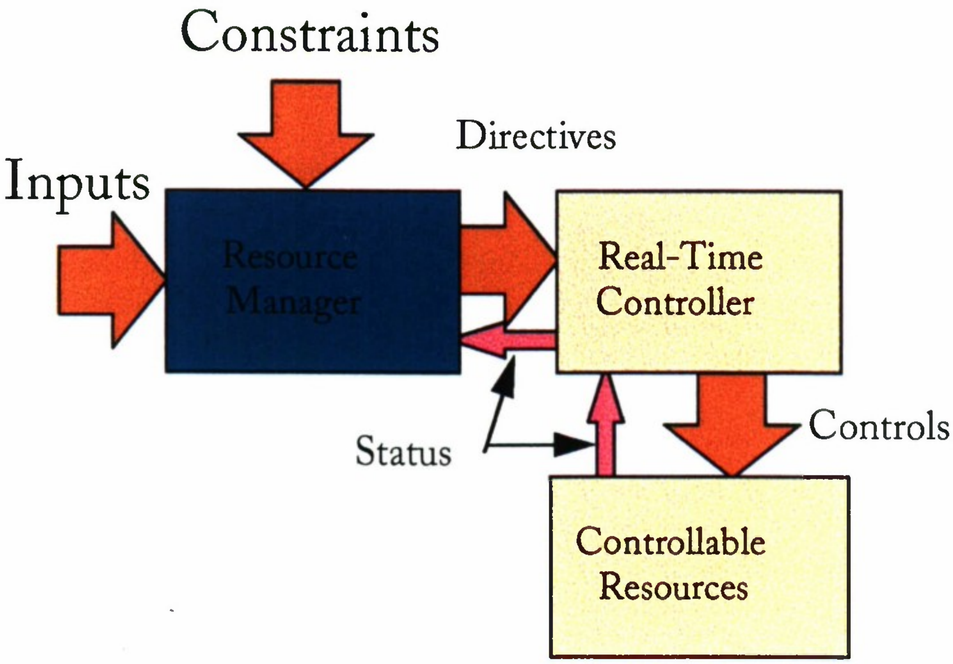


Figure 1. RF resource management and control processes.

1. Management Process, Inputs, and Constraints

Ultimately, the radio frequency (RF) management process for communications involves defining the network topology in real time. In the commercial environment of terrestrial and SATCOM links, this process exists only as long-term planning; there is no commercial counterpart to the real-time topology issues of a military multi-media network. The RF Network Management process operates under a number of constraints, and with multiple inputs. The communications output of the process is a "topology plan" for connectivity, including what frequencies, radios, antennas, cryptographic units and keys, etc., to use at each node—ship, aircraft, LCAC, armored personnel carrier, shelter, etc., down to the individual soldier or Marine. For Warfighter's Internet, this also includes planning the numbers, locations, and connectivites of the airborne nodes in addition to the radio system parameters.

The planning process must address the needs and concerns of all users of the spectrum. From the communications perspective, three key resource types are available for apportionment: spectrum, nodal equipment, and relays. However, the spectrum is a shared resource—its monitoring and apportionment properly are functions of the Command Control Warfare Commander (C2WC) [Navy] or equivalent at the various command levels of unit, force, and theater. Since proper communications planning and management must consider all three resource types, and since spectrum control is properly a C2WC function demanding inputs from, and providing direction to, multiple warfare support disciplines, the part of the communications planning and management called RF Network Management is a function that is best executed at the C2WC level.

The following paragraphs focus on inputs and constraints for RF network planning and management. Many of these discussions also apply to other spectrum users such as active surveillance.

1.1 User Demands (Real-Time Input)

The user needs of the moment (and when possible, projected forward in time) are an important input to the management process. These needs include: the desired connectivities; the offered loads at various nodes; priorities of the various traffic elements; and the formats of the traffic (voice, video, data). The needs will also normally include some type of Quality of Service (QoS) parameters, such as tolerable delays or error content, and packet versus stream service implications. The responsibility of the manager is to assure that resources are apportioned to highest priority users first, and that as many as possible of the demands are met. This could mean, for example, establishing a high capacity point-to-point connection to meet a temporary high-volume user need. User demands at concentrator nodes will fluctuate, but in general some level of service will be required at all times. User demands for individual users (such as handheld terminals) will generally be very low duty cycle. While statistical in nature, the individual user demands may be geographically correlated and tied to local operations (combat conditions). Repositioning of relay nodes might be one response to user demand.

1.2 Equipment Status and Performance (Constraint and Real-Time Input)

The management process will need to know the status of equipment at each location within the managed domain. Some factors, such as supportable data rate or frequency tuning range, will be fixed in most cases, while others, such as equipment failure or degradation, will vary with time. For limited-duration deployable relays (including "aircraft of opportunity" and dedicated relay UAVs), the manager will need to know the host platform operating limitations such as launch locations, flight profiles, operating ceiling, and endurance, as well as the number of units available and their cycle times.

1.3 Electromagnetic Compatibility (Constraint)

The management process must consider the impacts among multiple radiating systems on a single platform, or among multiple platforms in close proximity. The manager must select operating frequencies to maximize overall performance of all systems (not just communications systems) as a group. This implies identification of suitable performance measures for each of the systems. The Electromagnetics Systems Manager must do this for all communications systems at a given location, but the impact of electromagnetic compatibility extends beyond communications. Interactions with IFF, navigation, ISR (intelligence, surveillance, and reconnaissance), EW/IW, and perhaps other systems must also be considered.

1.4 Terrain (Constraint)

Terrain is a primary constraint on coverage from airborne nodes, and on connectivity for surface-surface communications. Blockage, diffraction, and reflections determine link performance in the presence of mountains, hills, canyons, and tall buildings (urban canyons). Vegetation penetration is an important consideration.

1.5 Operations Areas/Force Distributions (Real-Time Input)

The distribution of force elements is an important input for determining required coverage. Different force elements (individual users, concentrator nodes) will require different connectivity (duration and volume) and will be served by different equipment having different parameters.

1.6 Spectrum Availability (Constraint)

For radio frequency network management at the Unit, Force, and Theater levels, spectrum availability is considered an input provided by external sources. Frequencies available to DoD are negotiated at the national level. These are in turn assigned to the Services, and then divided up among forces. The apportionment among the Services should occur at the lowest level responsible for a Joint operation. The Electromagnetic Systems Management function then does the apportionment at progressively lower levels. Two types of inputs are possible: one indicates which frequencies to use, while another indicates which frequencies to avoid using. The latter input might also be in terms of avoiding interference with certain systems (belonging to another service, other countries, etc.).

1.7 Environment (Real-Time Input)

Two types of inputs are important for assessing link performance potential. The first consists of those factors that determine path loss and multipath propagation. These include solar activity and resulting ion density profiles, 0° isotherm height, precipitation, fog, atmospheric temperature inversions, and sea state. The second consists of the noise environment(s) at the intended receiver(s). This includes both natural and man-made noise, where man-made "noise" may be other users of the spectrum, incidental noise (such as from nearby welding operations), or intentional hostile jamming. Some of these inputs can be anticipated either statistically (solar activity; rain storm movement) or operationally (scheduled maintenance welding operations), while others rely almost exclusively on real-time inputs (other spectrum users; hostile jamming).

1.8 Node Locations and Dynamics (Real-Time Input)

This input applies primarily to backbone nodes and concentrator nodes. Most of these network participants are mobile. For SATCOM using spot beams, this may require beam pointing to track units, or switching between beams as a mobile unit moves between spots. Communications via line-of-sight links requires that the units be within line of sight. For HF surface wave communications, there may be a power allocation or data rate selection that depends upon separation between the units, among other factors. Node locations are the single most important dynamic factor determining "best" or "good" network topology to meet user demands. Node locations are generally available in real time from combat direction systems. For airborne "relay nodes of opportunity," the aircraft type, flight profile, and capability as a relay must be considered in planning the network topology evolution.

1.9 Policy, Doctrine, and Real-Time External Controls (Constraint or Real-Time Input)

Policy and doctrine are longer-term constraints that generally apply over months or years. These might, for example, be based on the need to support a command hierarchy. These could include assignment of DSCS SATCOM resources, or the definition of the controlling authority for a particular resource. Real-time external controls might be, for example, usurpation of a SATCOM channel for State Department use during a crisis. Other real-time controls in this category include imposition of EMCON or partial EMCON, whether generated at the node performing Electromagnetic Systems Management or elsewhere. Another example of a real-time control falling in the policy area would be restrictions on emissions in accordance with HERO/HERP/HERF (hazardous effects of radiation on ordinance, personnel, and fuels) conditions.

1.10 Threat Locations (Real-Time Input)

Threats to the Warfighter's Internet include physical destruction of backbone nodes (e.g., air-air or surface-air missiles against airborne relay nodes, artillery against ground relay nodes), jamming, and intercept/localization of ground/sea forces. Known and suspected threat locations must be considered when planning backbone node locations.

2. Management Configuration Concepts

The functional relationships in an Electromagnetic Systems Manager are shown in Figure 2, which is a Navy shipboard example. An implementation for any joint command node would be functionally similar. Note that the figure shows only the relationships associated with the Electromagnetic Systems Manager; relationships among the other functions on the diagram (such as between the routing and switching function and the radio systems) are not shown. This does not at all imply that such relationships do not exist.

In this shipboard example, the planning and management function (and its associated database), would reside on a GCCS machine supporting the C2WC (or organizational equivalent at the unit or theater level. The control (as opposed to management) function for JMCOMS resides in the Advanced Digital Network System (ADNS). From the communications perspective, the primary output of the Electromagnetic Systems Manager is the network configuration definition. As used here, network configuration is intended to include the following items:

- The endpoints (nodes) of all point-to-point links.
- The frequencies, antennas, transmit power level, modulation types, data rates, coding, multiplexing formats, crypto types and keys, and other critical parameters for all point-to-point links.
- The node membership list for all broadcast-mode networks.
- The frequencies, antennas, transmit power level, modulation types, data rates, coding, multiplexing formats, crypto types and keys, and other critical parameters for all broadcast nets.

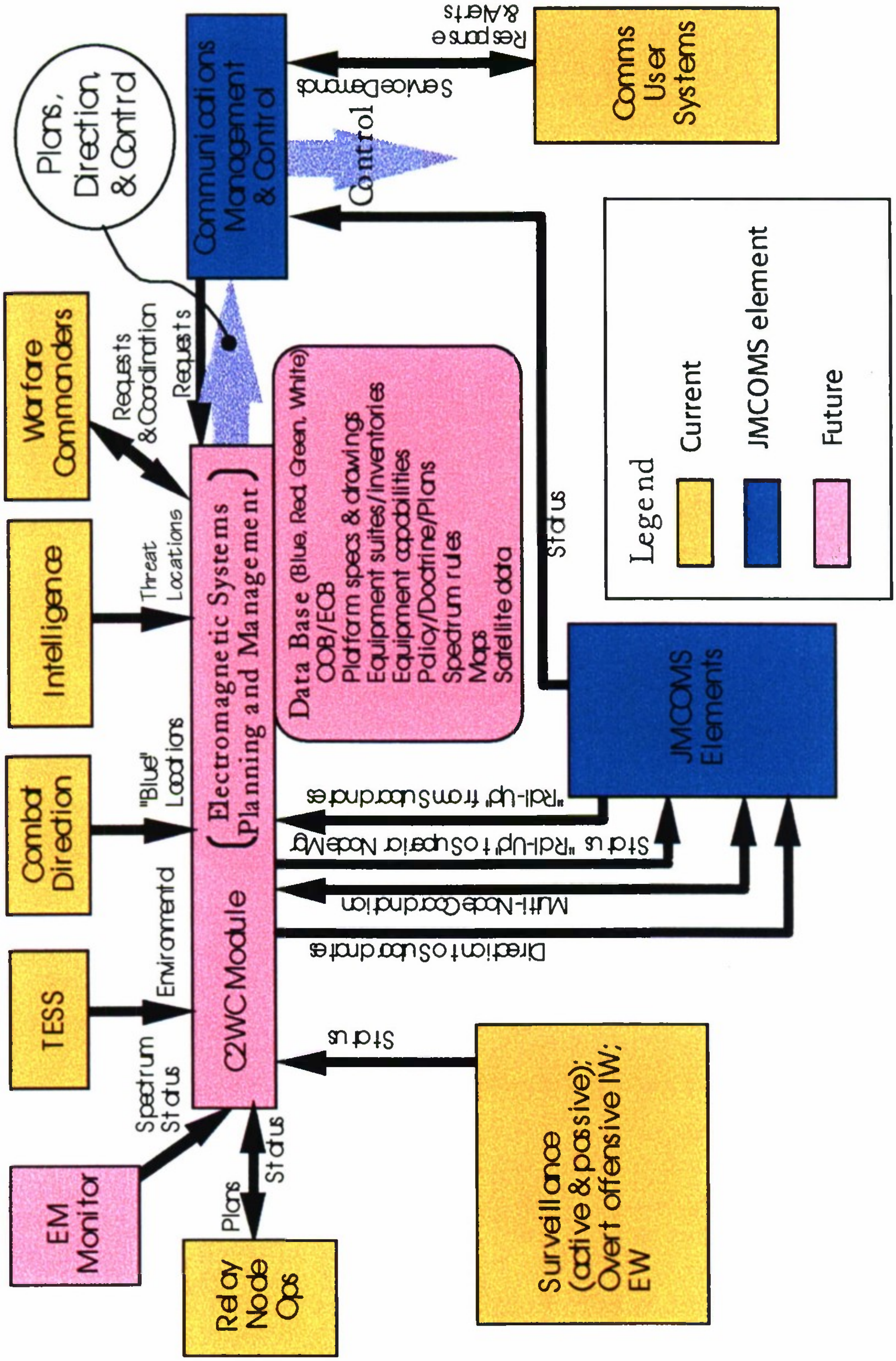


Figure 2. Functional relationships for an Electromagnetic System Manager (Navy Shipboard Example).

The network configuration is used by the ADNS (or equivalent) controller to determine how to set up nodal equipment to comply. On Figure 2, the output from the Communications Management and Control function is labeled "Control." This output is not the full configuration for the network, but only the subset pertinent to the local node.

Electromagnetic Systems Management relative to communications is envisioned as a distributed, hierarchical process. Each node computes the network configuration for itself and its subordinates. The manager uses communications to coordinate the process. Information on status (such as key equipment failures or significant deviation from expected user demand) gets passed both up and across ("coordination") the hierarchy. Decisions are passed down the hierarchy only if they differ from the anticipated decision. That is, each node knows the information available to its subordinates, and their algorithms. Each node then runs the process the subordinate runs, and compares the solution with its own solution. Significant differences are reported to the subordinate(s).

The figure is intended to cover all inputs and outputs of the process across all levels, from Unit, through Force, to Theater. Not all nodes will have all of the functionality. For example, the relay node operations function would only exist at locations where relay nodes are controlled. To minimize loading on the communications channels by the planning process, all nodes would include most (perhaps all) of the planning algorithm functionality. At nodes lacking a relay operations element, the relay node operations status would be algorithmically generated based on factors such as sustainable operations tempos, current relay locations (from CDS data) and past location history, policy, and current force operations or readiness level. When a node collocated with a relay operations center receives real status information that results in a network plan that differs from what would be generated using algorithmic inputs, the "delta" in status is distributed.

The resources labeled "Future" on the figure do not presently exist in the form required for the Electromagnetic Systems Management process. Each of these is briefly discussed in the following paragraphs.

C2WC Module. This is the software-based system to support the Command Control Warfare Commander. The C2WC is responsible for management and control of electromagnetic emission and reception. In support of this role, we view that the Electromagnetic Systems Manager, operated for and by the C2WC, would plan the communications network based on all available input data, plus guidance or constraints provided by higher authority (not shown on the figure). The planning and management processes would be conducted for all electromagnetic systems simultaneously. This is necessary to allow effective operation of all systems requiring access to the electromagnetic spectrum.

EM Monitor. This monitor senses the shipboard (or other local) electromagnetic environment, and provides this information to the Electromagnetic Systems Manager. This is envisioned as a product like the fiber-optic probe antenna investigated under ONR exploratory development funding. The monitor detects energy across the entire electromagnetic spectrum of interest to DoD electromagnetic systems planning, and provides the power spectrum information to the Electromagnetic Systems Manager.

The resources labeled "Current" on Figure 2 currently exist, and can provide the needed information for Electromagnetic Systems Management. These are briefly described in the following paragraphs.

Relay Node Operations Unit. This function includes the piloting functions for RPVs, UAV maintenance, relay pod operations, etc. Thus it is not a single functional unit, but a collection of functions implemented to different degrees at different nodes. For operating large land-based UAVs, this may be an organization of several full-time support people, with associated facilities. For other relay capabilities, it might be several people who perform relay-related functions only occasionally.

Communications User Systems. Users, via user systems, provide information concerning service demands, including capacity, duration, and quality of service. The Electromagnetic Systems Manager provides information to the user concerning the status of service. This may include: estimates of when a large file will be transferred, start and end times for a planned video conference, alerts to planned service downgrade to a user (caused, for example, by temporary peak loads), and explanation for current service degradation (e.g., service termination in favor of higher priority users) together with estimates of restoral time. Note that the user systems exist today, but appropriate manager interaction and interaction techniques do not. The focus of this paragraph is on communications users, but other users (e.g., surveillance systems/operators) also present operating demands to the Electromagnetic Systems Manager.

Tactical Environmental Support System (TESS). TESS provides the weather data required to support Electromagnetic Systems Management.

Combat Direction. The combat direction system has current information on the positions, movement, and intended movement of forces.

Intelligence. Intelligence sources can provide estimated threat locations needed to support Electromagnetic Systems Management. Locations of suspected hostile intercept sites for direction-finding and localization are useful when attempting to plan low probability of intercept communications. Locations of potential jammers, and their anticipated intent, are useful for planning anti-jam communications. Alerts concerning other possible activities (allied, hostile, neutral) that might affect communications are also provided, including information on potential physical threats to relay nodes.

Non-Communications Electromagnetic Systems. These systems include both active and passive electromagnetic surveillance systems, navigation systems, identification friend or foe systems, and electronic warfare systems. The status of each of these systems is provided as a planning and management input. Status includes faults/failures, and current operating mode(s).

The systems labeled "JMCOMS element" on Figure 2 are planned parts of the JMCOMS shipboard implementation. The following paragraphs briefly discuss these. Other concentrator nodes, such as RAPs, would provide similar functionality.

JMCOMS Elements. This includes routers, switches, multiplexors, and radio system components. Elements provide information on: current connectivity, both internal to a node and external in the form of equipment connection paths; and the status of equipment, including how it's connected and configured as well as the results of equipment built-in tests. These elements also provide the communications links used by the Electromagnetic Systems Manager to pass status updates and network configuration plans, and to obtain external guidance.

Communications Management and Control. This element provides the detailed implementation to configure the communications equipment in response to the current management direction provided by the Electromagnetic Systems Manager. This element also provides all of the traditional network management functions.

Each large tactical unit would have an Electromagnetic Systems Manager. Each would also have downloaded data and any necessary models, relevant to every unit. Each unit would have a local network of heterogeneous computing resources to do their own calculations as required. Very small units might have only a subset of the total capability, preferably with no operator required.

Thus each tactical unit, having virtually all of the necessary data and models for representing itself, the environment and every other unit, could independently calculate the RF asset configurations and schedule for itself, its subordinate nodes, and its peers. The calculated results at each unit would be virtually identical to those calculated by every other unit. This will provide the basis to allow a network to be initiated and then adapted in real-time to support required services even in the most hostile environments. Once initiated, the network operation can be refined as a result of data exchange among the units. Real-time adaptations would be required for a variety of changes, e.g., heading changes by tactical units result in antenna radiation pattern nulls being pointed in different directions, or radar coverages change as units maneuver.

The information to be processed by the Electromagnetic Systems Manager spans a variety of classification levels, from unclassified for various equipment parameters, to highly sensitive for own-force locations, intelligence inputs, and user service demands. The processing algorithms will need to operate at a system high level corresponding to the level of the most sensitive inputs. Outputs generated by the process may be at various levels which have not yet been identified. This may result in the need for a multi-level secure environment.

3. RF Network Management: Warfighter's Internet

From the discussion of the previous section, it should be clear that current commercial products address management of network elements (e.g., routers) in the ISO Management Framework sense, but are not intended to address many of the military management needs in the Information Reference Model for DoD Network Management Systems sense. Even for the earliest WI demonstration (FY-99), some additional capabilities, such as flight profile planning, will be required. It should be possible to perform most of the planning functions manually for the demonstrations, since the networks will be small. Eventually, tools to support the processes outlined in Figure 2 will be required.

For a network of a small number of nodes in a relatively stable configuration, such as that shown in Figure 3, the backbone connections can be planned manually (as was done for the figure). Such a plan is adequate for a fixed geometry with fixed coverage, and with tight-orbiting air nodes or fixed (e.g., mountain-top) ground nodes. (The configuration shown is for a case where each backbone node can support no more than two backbone links, and where a closed path is desired to assure full connectivity in a case where any single link fails.) However, manual planning is inadequate when the topology is changing rapidly, as with a moving "relay of opportunity" aircraft, or where the supported forces are changing their distribution on the ground.

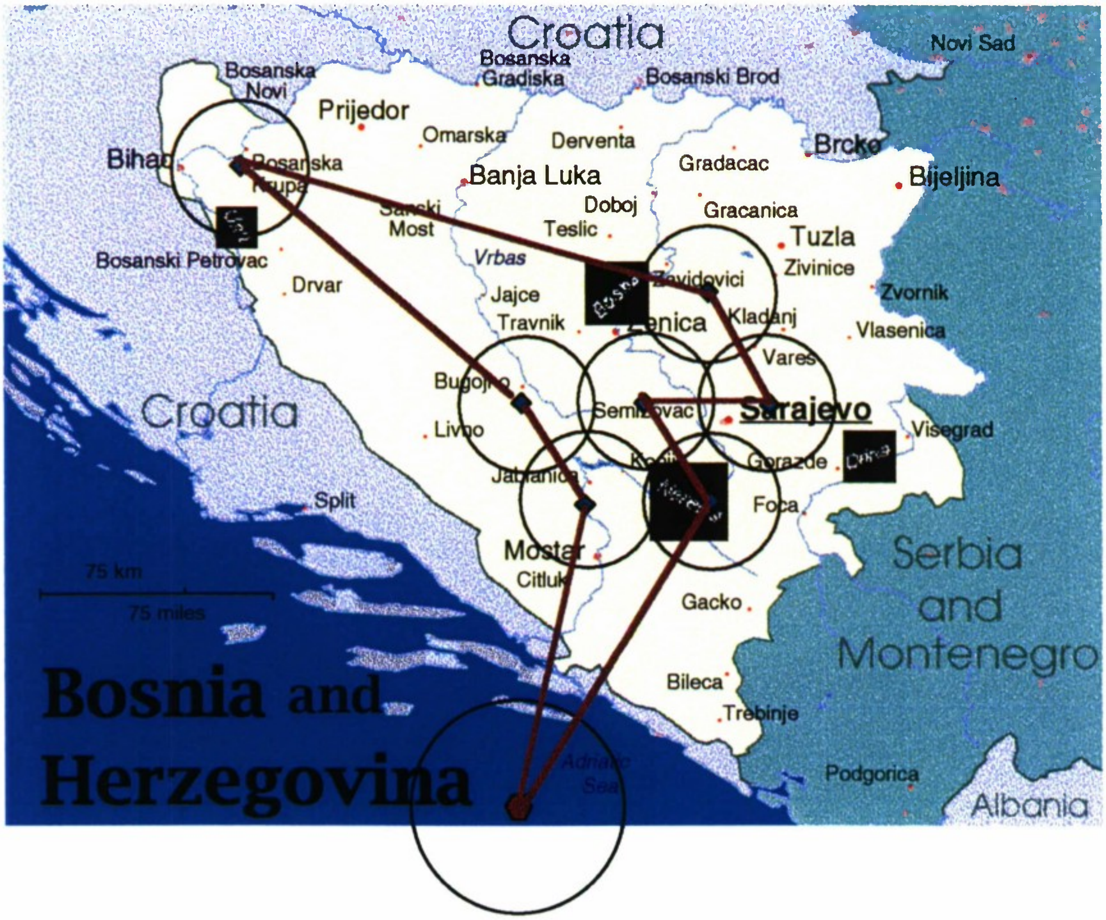


Figure 3. Example manually constructed backbone topology.

The distribution shown in Figure 3 is intended to suggest scale rather than to represent any real case. The circles around the nodes represent ground coverage for airborne nodes operating at an altitude of 20 km, with smaller circles for 30° minimum look angle, and the larger circle for 20° minimum look angle. Over the sea, an even smaller minimum look angle should be supportable. The different color/shape of the relays is not significant.

The process for finding an "optimum" path connecting the nodes is related to the classic traveling salesman problem. An algorithm for finding the optimum solution, other than exhaustive search, does not exist. For small networks like the WI backbone, exhaustive search can be used. For larger numbers of nodes, algorithms exist for finding "good" solutions from among the many possibilities[3].

[3] For a closed-loop network connecting n nodes, there are (n-1)!/2 possibilities. This number grows very rapidly with n. For 6 nodes, there are 60 possibilities; for 11 nodes, there are 1814400 possibilities; for 20 nodes there are 6 x 10^{16} possibilities.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave blank*)

2. REPORT DATE
28 January 1998

3. REPORT TYPE AND DATES COVERED

4. TITLE AND SUBTITLE
Architecture and Concept of Operations for a
Warfighter's Internet, Volume 2: Appendices

5. FUNDING NUMBERS
C — F19628-95-C-0002
PR — 602

6. AUTHOR(S)
Edited by MIT Lincoln Laboratory

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Lincoln Laboratory, MIT
244 Wood Street
Lexington, MA 02173-9108

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
DARPA/ISO
3701 N. Fairfax Dr.
Arlington, VA 22203-1714

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
ESC-TR-97-065

11. SUPPLEMENTARY NOTES
None

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

This report consists of two volumes. This is Volume 2: Appendices, which contains Appendices A–L of the main report (Volume 1).

Military operations in the twenty-first century will he conducted in an increasingly information-rich environment. But delivery of this information is difficult in the forward areas of the tactical theater where current communications equipment is slow to deploy and not matched to the mobility of forward forces. This report describes a "Warfighter's Internet" that provides responsive data and voice communications to individual warfighters using hand-held cellular-like handsets that have a wireless connection to a hackbone network of airborne communications nodes that are within line of sight of theater forces, within line of sight of each other, and are also connected to command and support centers through the Global Grid. The airborne nodes self-deploy and the handsets arrive with the users, so the network can he available immediately. This report describes design approaches to meeting the technical challenges in this system, which lie in the adaptation of Internet data protocols to the mobile environment (commercial cellular systems do not have to deal with both mobile users and a mobile backbone) and to the efficient use of wireless capacity for hursty data transmission.

14. SUBJECT TERMS
wireless communications
data networking
data protocols
tactical communications

15. NUMBER OF PAGES
205

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT
Unclassified

18. SECURITY CLASSIFICATION OF THIS PAGE
Same as Report

19. SECURITY CLASSIFICATION OF ABSTRACT
Same as Report

20. LIMITATION OF ABSTRACT
Same as Report